Autonomous Adaptive Acoustic Relay Positioning

by

Mei Yi Cheung

B.A., Columbia University, 2011

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2013

Author . . . .
Department of Mechanical Engineering
August 15, 2013

Certified by . . . .
Franz S. Hover
Finmeccanica Career Development Professor of Engineering
Thesis Supervisor

Accepted by . . . .
David E. Hardt
Chairman, Department Committee on Graduate Students

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2013**	2. REPORT TYPE	3. DATES COVERED **00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE
Autonomous Adaptive Acoustic Relay Positioning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology,Department of Mechanical Engineering,Cambridge,MA,02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We consider the problem of maximizing underwater acoustic data transmission by adaptively positioning an autonomous mobile relay so as to learn and exploit spatial variations in channel performance. The acoustic channel is the main practical method of underwater wireless communication and improving channel throughput and reliability is key to improving the capabilities of underwater vehicles. Predicting the performance of the acoustic channel in the shallow-water environment is challenging and usually requires extensive modeling of the environment. However, a mobile relay can learn about the unknown channel as it transmits. The relay must balance searching unknown sites to gain more information, which may pay off in the future, and exploiting already-visited sites for immediate reward. This is a classic exploration vs. exploitation problem that is well-described by a multi-armed bandit formulation with an elegant solution in the form of Gittins indices. For an autonomous ocean vehicle traveling between distant waypoints, however, switching costs are significant. The multi-armed bandit with switching costs has no optimal index policy, so we have developed an adaptation of the Gittins index rule with limited policy enumeration and asymptotic performance bounds. We describe extensive shallow-water field experiments conducted in the Charles River (Boston, MA) with autonomous surface vehicles and acoustic modems, and use the field data to assess performance of the MAB decision policies and comparable heuristics. We find the switching-costs-aware algorithm offers superior real-time performance in decision-making and efficient learning of the unknown field.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**79**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Autonomous Adaptive Acoustic Relay Positioning

by

Mei Yi Cheung

Submitted to the Department of Mechanical Engineering
on August 15, 2013, in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

We consider the problem of maximizing underwater acoustic data transmission by adaptively positioning an autonomous mobile relay so as to learn and exploit spatial variations in channel performance. The acoustic channel is the main practical method of underwater wireless communication and improving channel throughput and reliability is key to improving the capabilities of underwater vehicles. Predicting the performance of the acoustic channel in the shallow-water environment is challenging and usually requires extensive modeling of the environment. However, a mobile relay can learn about the unknown channel as it transmits. The relay must balance searching unknown sites to gain more information, which may pay off in the future, and exploiting already-visited sites for immediate reward. This is a classic exploration vs. exploitation problem that is well-described by a multi-armed bandit formulation with an elegant solution in the form of Gittins indices. For an autonomous ocean vehicle traveling between distant waypoints, however, switching costs are significant. The multi-armed bandit with switching costs has no optimal index policy, so we have developed an adaptation of the Gittins index rule with limited policy enumeration and asymptotic performance bounds. We describe extensive shallow-water field experiments conducted in the Charles River (Boston, MA) with autonomous surface vehicles and acoustic modems, and use the field data to assess performance of the MAB decision policies and comparable heuristics. We find the switching-costs-aware algorithm offers superior real-time performance in decision-making and efficient learning of the unknown field.

Thesis Supervisor: Franz S. Hover
Title: Finmeccanica Career Development Professor of Engineering

Acknowledgments

Firstly, I would like to thank my thesis advisor, Professor Franz Hover, whose guidance and insight made this project possible. I was first introduced to the challenging field of marine robotics when I joined his group two years ago and have learnt a great deal from him since. I would also like to thank all my labmates for their help and expertise, especially Eric Gilbertson, Josh Leighton and Brooks Reed. An autonomous kayak takes many people to launch and it has been an equally collaborative experience at Hovergroup. Finally, I thank my parents, I would not be here without their support, and my brother, who always encourages me to strive harder.

Thanks also to Toby Schneider and Mike Benjamin at MIT for all their help with using Goby and MOOS, and to Keenan Ball and Sandipa Singh at WHOI for their quick responses and help when we had questions about the Micro-modems. I also thank MIT Sailing Master Franny Charles and Gerard for their patience and understanding with all the marine robots careening about the MIT Sailing Pavilion.

This work is supported by the Office of Naval Research, Grant N00014-09-1-0700, the National Science Foundation, Contract CNS-1212597, and Finmeccanica.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The ocean is a challenging environment for humans and robots alike. Features of interest may vary over wide temporal and spatial scales, the environment is often unknown and harsh as pressures and temperatures reach extremes, and familiar robotic senses like sight and touch degrade underwater. Large advances in marine robotics and autonomy have been made in recent years as scientists and engineers have risen to the challenge, motivated by the immense potential for robotics to further study of the ocean and usage of its resources. As individual vehicles, their navigation, control and communications, mature into commercially-available technology, the next step lies in enhancing the control and behavior of fleets of vehicles, so that they may organize and collaborate in a way that enhances the capabilities of the group as a whole.

Effective underwater communication is vital to the development of multi-vehicle applications in the ocean, and the main mode of wireless data transmission underwater is acoustic. Electromagnetic waves are severely attenuated underwater, rendering radio frequency (RF) communications inappropriate and optical communications short range (on the order of a hundred meters). However, the acoustic channel is notoriously unreliable and highly dependent on the properties of the medium. Modern channel modeling, estimation and coding schemes go a long way towards improving channel performance (e.g. [46], [16], [40]), but cannot directly address the channel's dependence on spatially variable environmental properties — temperature, salinity, bathymetry, surface conditions etc., and the physical location of the acoustic nodes.

This thesis considers the problem of maximizing the performance of the acoustic channel in an unknown environment by deploying a mobile, autonomous acoustic relay that adaptively positions itself so as to exploit horizontal spatial variability in the channel. As the relay learns about the channel's properties by transmitting, this can be posed as a classic exploration vs. exploitation scenario with an elegant and practical solution.

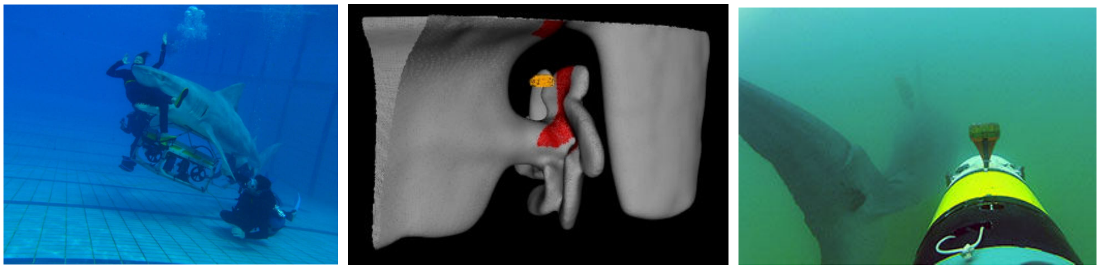
1.1 Background and Motivation

We briefly review common marine robotic platforms and applications, and the attendant technical challenges.

1.1.1 Marine Vehicles: Applications and Autonomy

Robotic ocean vehicles are primarily utilized by three general industries; oil and gas, defense and security, and scientific research (Fig. 1-1). In 2010, the oil industry represented 50% of global ROV sales and 20% of global AUV sales [11]. Development of vehicles capable of survey and inspection as well as equipment installation, manipulation and repair tasks has been motivated by dwindling shallow-water oil reserves and the move towards technically-challenging deep water wells [50]. In the security industry, robotic vehicles are well-suited to time-consuming tasks or dangerous tasks such as patrolling, reconnaissance, ship-hull inspections and mine counter-measures (MCM) (e.g. [4], [20]). Vehicles equipped with various sensors and able to communicate with static sensor networks are often used to collect oceanographic data on a large scale and in harsh environments. Increased availability of salinity, temperature, density, chemical and biological composition data help researchers better understand and model ocean processes [30]. Imaging, sonar or vision-based, allows vehicles to capture evidence of new marine species as well as uncover underwater archaeological sites [21].

Remotely-operated vehicles (ROVs) are designed for remote human control and use with a support vessel or platform. Power and commands are transmitted from

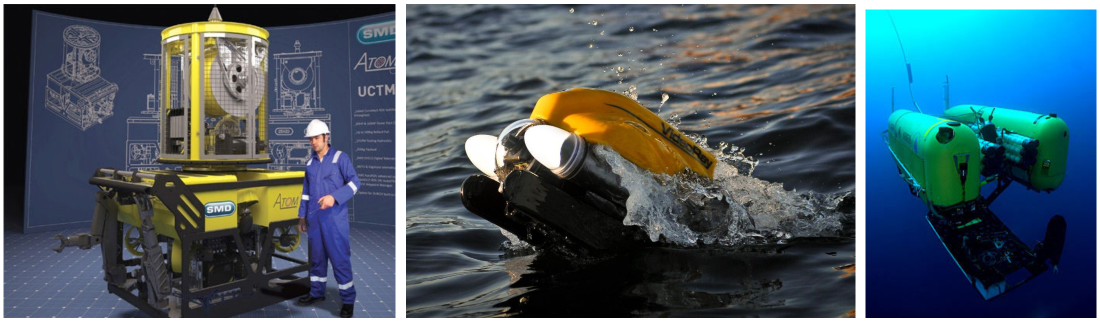


(a) Saab ROV *Falcon* (b) MIT Bluefin HAUV (c) Hydroid *REMUS*

Figure 1-1: Saab Seaeye's ROV *Falcon* controls a robotic shark to gently bite an actor; The Bluefin HAUV builds a point cloud mesh model of a ship hull to inspect for mines; Hydroid's AUV *REMUS* tracks a great white shark for Discovery Channel's Shark Week. Image Sources: (a) `http://www.rovworld.com/article2886.html`, (b) `http://web.mit.edu/hovergroup`, (c) `http://www.whoi.edu/oceanus/viewArticle.do?id=173392`

the support vessel to the vehicle through a long tether, while sensor data such as real-time video is transmitted back. Tethered ROVs operate underwater without the constraints of battery life and computing capacity but the presence of the long tether complicates vehicle dynamics and increases operating costs and complexity. High bandwidth data transmission through the tether enables real-time imaging, manipulation and pilot control, thus ROVs are generally used to replace divers for deep water equipment construction and repair tasks [51]. Commercial ROVs range from large, versatile work-class vehicles like Soil Machine Dynamics (SMD) *QUANTUM* and *ATOM*, to the specialized man-portable inspection and survey ROV VideoRay (Fig. 1-2). In general, ROVs offer real-time sensing, power and fine pilot control underwater while being relatively time, manpower and capital-intensive to operate.

However, as underwater vehicle technology matures and the price of vehicles decreases, human and ship support dominate as the major cost driver of underwater ocean operations. Autonomous underwater vehicles (AUVs) are able to operate independently of a support ship or platform for long periods of time (Fig.1-3). They have varying levels of autonomy, ranging from pre-planned lawnmower-style paths for survey and data collection to adaptive behaviors that respond to environmental stimulus. AUVs must handle their own low-level control and navigation, mission-level decision-making and operate sensor payloads for oceanographic data collection,



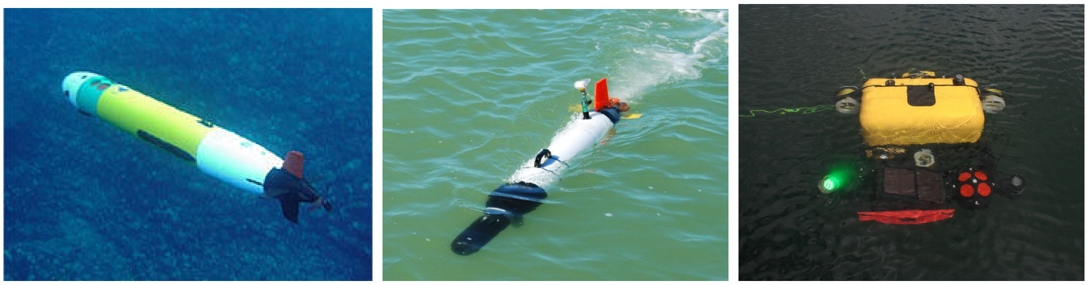
(a) SMD *ATOM* (b) *VideoRay* (c) WHOI *Nereus*

Figure 1-2: Work-class ROVs like SMD's *ATOM* are often used for construction and maintenance; ROV VideoRay is a small specialized robot for inspection and survey; WHOI's Hybrid Oceanographic Research ROV *Nereus* is able to operate in different modes. Image Sources: (a) `http://www.smd.co.uk/products/work-class-rovs/atom.htm`, (b) `http://www.videoray.com/homepage/professional-rovs/videoray-pro-4.html`, (c) `https://www.whoi.edu/main/nereus`

optical and sonar-based imaging. Currently, the majority of AUVs sold are rated for water depths less than two hundred meters, with the emphasis on small, light vehicles [11]. Their operational capability is limited by onboard computing power, battery life and means of actuation. Gliders are one type of AUV that move by adjusting the vehicle's buoyancy. Using fixed wings, they are able to travel in a vertical yo-yo pattern with very little power consumption. Hybrid AUVs such as the Bluefin HAUV and WHOI's *Nereus* are able to move autonomously underwater, but communicate with the support vessel through a thin, high-bandwidth fiber-optic cable.

1.1.2 Underwater Acoustic Communication

Wireless communication with underwater autonomous vehicles is key to increasing their ability and reliability, and the acoustic channel is the main practical carrier for wireless underwater data transmission over long distances. Radio frequency waves are heavily attenuated underwater (on the order of tens of meters), and high bandwidth optical communications is limited to short ranges (about one hundred meters in clear water). The acoustic channel is a wide-band packet-based erasure channel. Latency is high as the speed of sound in water is approximately 1500m/s and data



(a) Hydroid *REMUS* (b) OceanServer *Iver2* (c) MIT Bluefin HAUV

Figure 1-3: Hydroid's *REMUS* is well-suited to methodical surveying and mapping, OceanServer's *Iver2* is a low-cost, single man-portable system, and the Bluefin HAUV was designed for hull inspection with a high-resolution imaging sonar. Image Sources: (a) `http://www.whoi.edu/instruments/viewInstrument.do?id=1759`, (b) `http://iver-auv.com`, (c) `http://www.bluefinrobotics.com/products/hauv/`

rates are low. The channel is wide-band, in the sense that the bandwidth is not negligible with respect to the center frequency, and cannot accommodate multiple users or signal multiplexing easily. Numerous sources of signal interference make packet decoding a challenge, for example, Doppler effects from source and receiver motion, wave refraction due to the nonlinear sound-speed profile, and reflection and scattering from the surface, bottom and particles within the volume creating numerous multipath arrivals (Fig. 1-4). These sources may vary widely in time and space and are heavily dependent on local environmental properties including density, temperature, salinity of the water as well as bathymetry and surface conditions. In harbor and man-made environments, structures and ambient noise can be a problem. Given sufficient knowledge of the environment, ray and beam-tracing may be used to predict the effects of multipath as well as the location of "shadow-zones" and acoustic wave guides in which refraction of sound waves create zones where acoustic waves do not enter or exit respectively ([38],[10]). These methods are generally conducted in the two-dimensional vertical plane where the sound-speed profile is well-defined.

The signal's attenuation and spreading loss is frequency-dependent, as delay occurs over many milliseconds due to the low speed of sound in water. Background ambient noise from sources like wind, waves and shipping may be approximated as Gaussian but not white, while site-specific noise such as snapping shrimp found

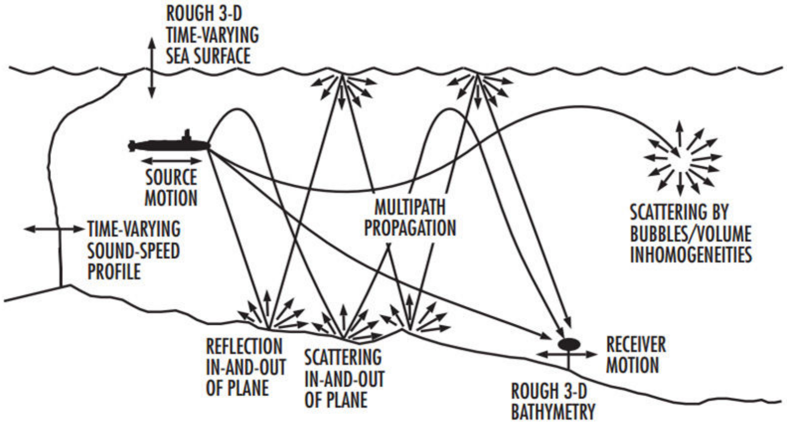


Figure 1-4: Illustration of possible sources of interference in the acoustic channel. Many of these sources are difficult to model and predict, especially in an environment where environmental properties may vary widely in time and space. Image Source: `http://www.rjeint.com/acousticTerms.htm`

only in certain areas of the world often contain significant non-Gaussian components ([46],[6]). Thus, the Signal-to-Noise ratio (SNR) of the channel is a function of the frequency and distance, and in particular, the available bandwidth decreases with transmission over increased distances. This means dividing long distances into multiple hops allows for transmission at higher data rates over each link and for lower total power consumption [46]. In recent years, development of advanced channel estimation and error correction schemes have gone a long way towards increasing the robustness of the point-to-point acoustic channel (see [45], [16] for recent surveys). Coherent phase-shift-keying (PSK) modulation schemes work well with sparse adaptive decision feedback equalizers (DFEs), and much work has been done on improving the real-time performance and robustness of the channel estimator and equalizer. A fundamental trade off exists between choosing to send more data in a larger packet, where error correction schemes can be more sophisticated with less overhead but the time taken to encode, transmit and decode the packet is longer and may incur more channel interference and packet loss, and sending a smaller packet for a lower overall data rate but more reliability in packet success. In general, acoustic modem parameters such as modulation type, error correction scheme, packet size and transmission power may be tuned heuristically to improve performance.

1.1.3 Formulation and Motivation

Many ocean applications are well-suited to the use of a team of vehicles collaborating and sharing information. Tracking and pursuit of dynamic ocean processes (e.g. fronts and plumes), marine animals and vehicles is a time-critical application in which multiple vehicles may contribute greater robustness, tracking precision and maneuverability (e.g. compared to long towed arrays). Oceanographic surveys require data collection over large time and spatial scales, and multiple vehicles may be able to resolve spatio-temporal ambiguity encountered by a single vehicle. As the availability of vehicles increases and human/ship support costs become proportionally more important, deploying additional vehicles in order to complete missions faster or multiple missions at the same time becomes increasingly cost efficient. The effectiveness of a team of underwater vehicles hinges on the performance of the acoustic channel for critical communications. In order to motivate the development of sophisticated behaviors and decision-making for groups of vehicles, we address the goal of improving the performance of the acoustic channel in an unknown environment, specifically, maximizing the cumulative data transmitted through the channel over the course of a mission.

For most missions in an unknown ocean environment, measuring the water properties, bathymetry and other environmental variables for detailed modeling of the acoustic channel is time-consuming and undesirable. Day-to-day fluctuations in surface and sea conditions as well as the large spatial scale of ocean applications make accurate modeling a challenge. In shallow-water and man-made environments, multipath interference results in significant variability in space, which may be exploited by acoustic nodes if the variations in channel performance was known. A key insight lies in recognizing that acoustic nodes can learn about the statistics of the channel (i.e. SNR, packet success rate) at their current location while receiving and transmitting. This is often exploited by static acoustic sensor networks with adaptive networking algorithms. For the purpose of improving acoustic communications during a multi-vehicle mission, we consider the case of deploying a mobile acoustic relay. The relay is free to travel to different locations and learn about the performance of the channel; i.e. the point-to-point physical channel in space. While the acoustic modems may correct for errors and interference within the channel itself, a mobile relay is able to adapt to link variations in physical space. Fig. 1-5 shows this setup in the Charles River, Boston MA. The mobile relay's goal is to maximize the cumulative transmissions from the source to the destination node, given no prior knowledge of the environment.

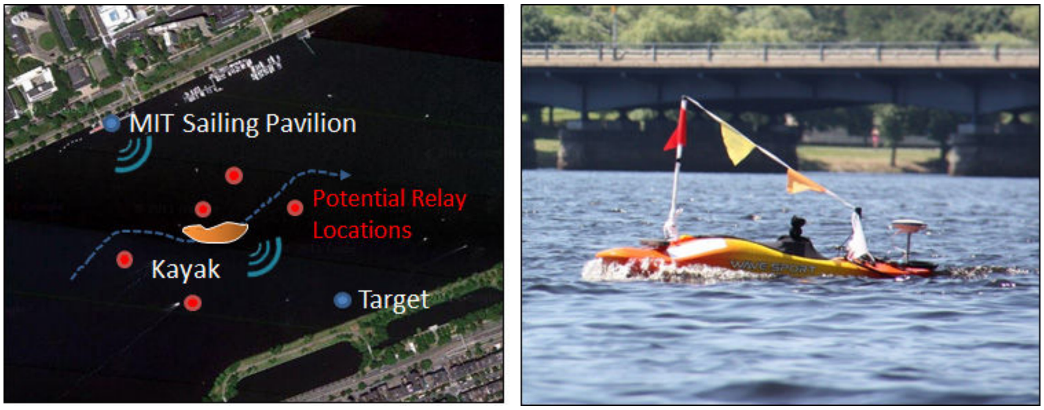


Figure 1-5: Illustration of adaptive relay positioning problem implemented in the Charles River, Boston MA with an autonomous surface vehicle (right) towing an underwater acoustic transducer as the mobile relay. A fixed source node is present at the MIT Sailing Pavilion and a fixed destination node is another kayak station-keeping across the river.

Thus, the relay must balance searching unknown sites to gain more information, which may pay off in the future, or exploiting already-visited sites for immediate reward. This is a classic exploration vs. exploitation problem that is well-described by a multi-armed bandit optimization framework. We formulate the multi-armed bandit for the problem of adaptive acoustic relay positioning and apply an elegant optimal solution in the form of Gittins indices. However, for an autonomous ocean vehicle traveling between distant waypoints, the time costs of traveling between locations (switching costs) are significant. The multi-armed bandit with switching costs has no optimal index policy, so we develop an adaptation of the Gittins index rule with limited policy enumeration and asymptotic performance bounds.

1.2 Prior Work

Though a conceptually simple problem, adaptive positioning in the spatial horizontal field so as to improve the performance of the acoustic channel has not been systematically studied before. Depth adjustment was studied recently by Detweiler et al. [18], following the discussion of Akyildiz et al. [3], with modems that were not COTS units. Packet success rate in water less than 10m deep showed variability by over a factor of two through the space, and with no clearly identifiable physical structure. Schneider and Schmidt integrate three acoustic modeling techniques, the Bellhop Ray Tracing model, OASES Wavenumber Integration Model and KRAKEN Normal Modes model with real-time CTD data in order to optimally adjust the depth of an AUV for maximum SNR [42]. Adaptive protocols for acoustic sensor networks have been developed to optimize over packet loss and network performance (e.g. [12], [26]), although efficient energy usage, routing and MAC protocols are more widely studied. For a recent survey, see Akyildiz [2].

Autonomy behaviors and algorithms have been developed for mission-level control and path-planning of multiple vehicle systems. Curcio *et al.* considered the use of multiple SCOUT ASVs for autonomous oceanographic survey with wireless internet links and present field experiments in Monterey Bay, CA and Dabob Bay, WA [17]. The vehicles collaborated to measure the sound speed in a section of water using ranging pings and conductivity-temperature-depth (CTD) measurements. The path-planning problem for adaptive sampling with single and multiple vehicles, in which the goal is to maximize the accuracy of the estimates was formulated as a mixed integer linear program (MILP) by Yilmaz *et al.* [53], in which the acoustic link was modeled as a distance constraint between the AUVs and the support ship. Munafo *et al.* consider a heuristic data-driven algorithm for the cooperation and coordination of a team of AUVs in an environmental mapping mission [34], in which the overall goal is to achieve a desired map accuracy. Each agent shares its information with the team, and the cooperation algorithm trades off remaining in communication with maximizing the local distance among the AUVs. The map estimation is based on radial basis functions

(RBFs), following the approach of Alvarez *et al.* [4]. A large-scale field experiment was reported by Leonard *et al.* [33] in which a fleet of six gliders was coordinated over twenty four days. A path-planning algorithm was used to fuse real-time data collected by the gliders with ocean model predictions in order to optimize sampling patterns and minimize the uncertainty of field estimates. Most recently, the MORPH (Marine robotic systems of self-organizing, logically linked physical nodes) project [37] conducted field trials at Toulon IFREMER site demonstrating co-operative path following and range-only formation control using teams of heterogeneous vehicles with wifi and acoustic communications.

Shankar and Chitre formulated the multi-armed bandit for tuning configurable parameters on acoustic modems, given only the bit error rate as estimated by a Kalman filter [43]. In robotics, Stone and Kraus [47] considered the formation of an ad-hoc team with varying ability and information (teacher and learner) as a bandit problem maximizing the reward over the team of agents. The robot grasping task has also been formulated as a multi-armed bandit by Kroemer *et al.* [29], in which a high-level hierarchical controller learns about the performance of various grasps using an Upper Confidence Bound (UCB) policy. The restless form of MAB with switching costs has been applied to the problem of task allocation and routing for UAVs [31], where a linear relaxation based on work by Bertsimas and Niño-Mora [9] computes the multi-agent route. Similarly, the linear relaxation solution was applied to relay selection optimizing over the physical layer in TCP wireless communications [49].

It has been shown that no optimal index policy solution exists for the MABSC [7], and most research on this topic has focused on deriving general properties of the optimal policy [5], deriving explicit optimal policies for special cases [19], and bounding approximations to the optimal policy [1]. For a recent survey, see Jun [27]. The problem has also been reformulated as a semi-Markov multi-armed restless bandit, addressed by marginal productivity indices (MPI) [36] and a linear programming relaxation (LP) [32], based on work by Bertsimas and Niño-Mora [9]. These include switching costs as a natural extension of the restless bandit [52], in which processes are non-stationary. However, both the MPI[1] and LP treatments of the restless bandit trade the advantage of an exact and general problem statement for a lookahead horizon limited to one switch/step. This may not be enough for some applications. The alternative — enumeration — incurs an exponential cost. Here, we exploit a small state space that allows for a much deeper enumeration, but also seek methods by which this load can be reduced. In particular, applying the key result of Asawa and Teneketzis [5] allows us to adapt the Gittins index policy while greatly reducing the computation cost of decision-making required.

1.3 Summary

The goal of this thesis is to address the problem of maximizing cumulative data transmission through the acoustic channel by adaptively positioning an autonomous and mobile acoustic relay. In Chapter 2, we formulate the multi-armed bandit optimization problem and its specific application to adaptive relay positioning with the inclusion of switching costs, and provide a solution in the form of an index-based decision policy. In Chapter 3, we describe extensive field experiments conducted with autonomous surface vehicles towing acoustic modems in the Charles River Basin, Boston MA. The performance of the acoustic channel in this complex, shallow-water environment is presented in Chapter 4, based on experimental data. In Chapter 5, we evaluate the effectiveness of the multi-armed bandit decision policy, autonomously, in comparison with competing algorithms and on synthetic data. This work has been published in [15], [13] and submitted to [14].

[1] For a stationary process with switching costs, the MPI is equivalent to Asawa & Teneketzis's switching index [36]

Chapter 2

Problem Formulation

2.1 The Canonical Multi-Armed Bandit

The canonical Multi-Armed Bandit is an optimization framework for resource allocation problems in which the resource must be allocated sequentially between a number of competing projects. The overall goal is to maximize the cumulative reward obtained by the resource, however, each allocation must trade off prioritizing immediate reward acquisition with taking action for potential future benefit (such as acquiring information). The name arises from the following gambling analogy:

A gambler (the *resource* to be allocated) can play one slot machine (or "one-armed *bandit*") at a time. He has to choose between multiple slot machines (or one slot machine with many arms, hence *"multi-armed bandit"*), and each arm returns a reward once played. The multi-armed bandit may be *deterministic*, each arm returning a reward from a fixed sequence, or *stochastic*, each arm returning a reward with a fixed probability distribution. In general, each arm is characterized by a different reward process, and the gambler begins with no knowledge of these processes. As the gambler plays the bandit and observes each reward obtained, he is able to update his *information state*, or his estimate of the reward distribution for each arm. Thus, the gambler learns from his actions and builds a model of the multi-armed bandit, using this model to improve future decisions. Each decision must balance improving his model through *exploration* (playing arms with poorly characterized distributions

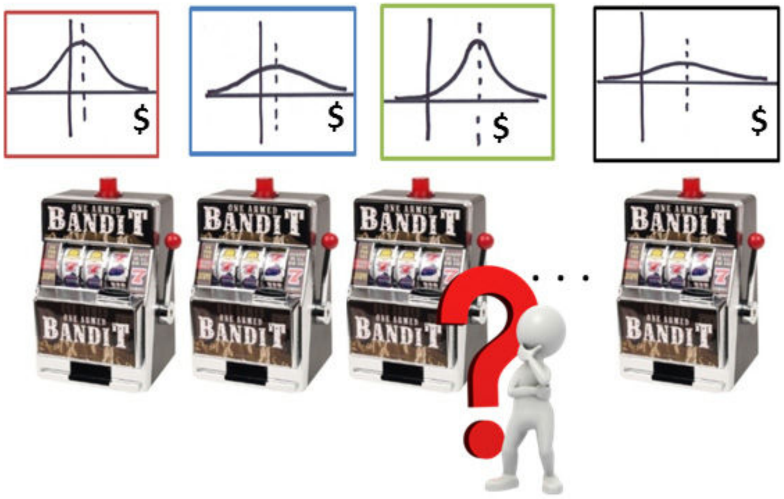


Figure 2-1: Illustration of multi-armed bandit gambling analogy. The gambler makes a series of sequential decisions between one-armed bandits (arms of the multi-armed bandit) with unknown reward distributions. Estimates of the unknown distributions are built up by the gambler over multiple observations, and inform his future decisions. Image Sources: `http://thetechnologicalcitizen.com/`, `http://www.thegadgetexperience.com/slot-machines/`

to gain more information), or *exploiting* the model to gain the greatest immediate reward (i.e. playing the arm with the current most favorable distribution). This fundamental tradeoff arises in many real-life scenarios, and the multi-armed bandit formulation can be widely applied to such problems as beam scheduling for array tracking systems [28], radio channel allocation [23], and website ranking [41].

2.1.1 Problem Formulation

The multi-armed bandit is in general a discrete-time Partially Observed Markov Decision Process (POMDPs), or decision process on a Hidden Markov Model (HMM), in which the underlying arm states are not directly observed and the observations (reward) are a probabilistic function of the unobserved Markov process. The stochastic one-armed bandit is defined as a sequence of process states $x(1), \cdots, x(n)$, where $x(n)$ is a random variable representing the state of the machine after it has been operated n times. The reward $R(x(n))$ from the state is a real, non-negative random variable.

The multi-armed bandit process is a collection of N independent one-armed bandit machines, indexed by i. The state of the multi-arm process as a whole is denoted by the vector $\bar{x}(t)$, containing $\{x_1(t) \cdots x_N(t)\}$. We denote the number of times machine i has been operated by n_i, and its state by $x_i(t)$, where t is the current global decision epoch:

$$t = \sum_{i=1}^{N} n_i. \tag{2.1}$$

In general, the underlying state space of the multi-armed bandit is exponential in the number of arms, rendering solving for the optimal solution computationally intractable (exponential in memory and computation). However, the curse of dimensionality can be addressed by assuming several key characteristics of the problem structure:

1. Only one arm is played at each time step (*decision epoch*)
2. Only the arm that is played returns a reward
3. Idle arms are frozen — i.e. arms that are not played do not change state
4. Switching the arm to be played is instantaneous and costless

Thus, at each decision epoch, the decision process samples a single machine, updating the state and reaping the associated reward, while the states of all other machines remain frozen. The optimal solution to this canonical formulation is a dynamic allocation policy, denoted by π, that defines at each decision epoch the machine for allocation i_t, such that the expected value of the total reward V_π is maximized. For the discount factor $0 < \beta < 1$ and an infinite horizon, this reward is:

$$V_\pi(\bar{x}) = E\Bigg[\sum_{k=0}^{\infty} \beta^k R(x_{i_k}(k)) \mid \bar{x}(0) = \bar{x}\Bigg]. \tag{2.2}$$

2.1.2 An Optimal Solution: Gittins Indices

Gittins and Jones [24] showed that the optimal policy is to play the machine with the largest expected reward per unit time, maximized over all stopping times $\tau > 1$: $i_{t+1} = \underset{i}{\operatorname{argmax}}(\nu_i(x_i(t)))$, where[1]

$$\nu_i(x_i(t)) = \max_{\tau > 1} \frac{E\left[\sum_{k=0}^{\tau-1} \beta^k R(x_i(k)) \mid x_i(0) = x_i(t) \right]}{E\left[\sum_{k=0}^{\tau-1} \beta^k \mid x_i(0) = x_i(t) \right]}. \tag{2.3}$$

Crucially, the index ν_i is a function only of $x_i(t)$, allowing the MAB to be decomposed into N independent stopping time problems. Various algorithms to calculate the Gittins index have been reported, recently by Sonin [44] and Niño-Mora [35].

2.2 A Heuristic Adaptation for Switching Costs

We define constant costs $c(i, j)$ to reflect the undesirability of switching from machine i to machine j; in the context of relay positioning, the cost is that of time spent in transit. If $t_v(i, j)$ is the time taken to travel from i to j, and $t_r(i, j)$ is the time taken to relay, we can set $c(i, j) = \lfloor t_v(i, j)/t_r(i, j) \rfloor$ — the number of transmissions the relay could have made on location if it had chosen to sample instead of traveling. This is only one of many cost models relevant to the application, and later we investigate several of them. The optimal solution to the MABSC is one that maximizes:

$$V_\pi(\bar{x}) = E\left\{ \sum_{k=0}^{\infty} \beta^t \left[R(x_{i_k}(k)) - \ c(i_k, i_{k-1}) \right] \mid \bar{x}(0) = \bar{x} \right\} \tag{2.4}$$

where we define $i_{-1} = i_0$. As noted previously, switching costs do not admit an index policy [7] because the reward returned by a process no longer depends solely on the number of times n_i an arm has been operated. For this problem, we describe a solution of the *priority-index policy* form, where separate "continuation" and "decision"

[1] This standard notation directly shows the form of expected discounted reward over discounted time, although in our formulation we assume β to be constant and independent of state.

indices are used [36]. This scheme separates the decision process into two modes. At every decision epoch, the continuation index is computed to decide if the current arm is continued. If it is not, the decision index is then computed to decide which arm to switch to. The continuation index ν_i is taken to be the Gittins index previously defined. If the current arm has the highest Gittins index of the field, it can be continued without further decision. However, even if it is not the current maximum, Asawa and Teneketzis showed that it is optimal to continue playing an arm up to its stopping time τ, only making a decision to switch when the stopping time is achieved (A&T Thm. 2.1) [5]. This occurs when the Gittins index of the current arm falls below any value it has previously reached, thereby defining the continuation rule:

$$\text{if } \min_{k<t} \nu_{i_k}(x_{i_k}(k)) \leq \nu_i(x_i(t)), \text{ set } i_{t+1} = i_t. \tag{2.5}$$

The continuation rule can only increase the number of times an arm is played. When the stopping time is achieved, i.e., the above condition does not hold, the decision index determines which arm to switch to. The continuation rule reduces the required computation frequency of the decision index, admitting an accurate and flexible but computationally intensive solution for a problem of this scope. We calculate the decision index by maximizing an m-horizon look-ahead enumeration of the expected reward rate over all possible policies π, where π is any possible sequence of plays $i_1, ..., i_m \; \forall i \in 1, ..., N$. We do not enumerate the action of remaining in the current location, although policies include choosing to return to the current location after switching away. The value of being in the final state $\hat{x}_i$ is accounted for with an updated Gittins index ν_i for that policy. Location-based switching costs are simple to include in this formulation:

$$\eta_\pi(\bar{x}(t)) = \frac{e(\bar{x}(t))}{E\left[\sum_{k=0}^{m} \beta^k \mid \bar{x}(0) = \bar{x}(t)\right]} + \nu_i(\hat{x}_{i_{t+m}}(t+m)), \tag{2.6}$$

where

$$e(\bar{x}(t)) = E\Bigg\{ \sum_{k=0}^{m} \beta^k \big[R(x_{i_k}(k)) - c(x_{i_k}(k), x_{i_{k+1}}(k+1))\big] \mid \bar{x}(0) = \bar{x}(t) \Bigg\}. \tag{2.7}$$

The adapted decision rule for MABSC is then

$$i_{t+1} = \underset{i_1}{\mathrm{argmax}}(\eta_\pi(\bar{x}(t))). \tag{2.8}$$

If $m = 0$, this rule is identical to the MAB Gittins index rule. If $m = 1$, this rule is identical to the switching index defined by Asawa. Since enumeration is computation-intensive, we apply A&T Thm. 2.1 to reduce the number of required decision index computations. Thus, longer horizons can be enumerated, allowing the algorithm to capture the benefits of efficient routing where a more myopic policy would not. An algorithm for enumeration is presented in Appendix C.

2.3 Bernoulli and Normal Reward Processes for Adaptive Relay Positioning

For learning and decision-making by the mobile acoustic relay within the MAB framework, we discretize the physical space into N potential relay locations and define each location as an independent arm of the bandit. In general, these relay locations may be dictated by mission constraints. We note that although programmable modem parameters such as packet encoding scheme can be included combinatorially as additional machines, we have fixed these for simplicity. The agent plays an arm by relaying through that location, updating its state information on the arm, and then deciding which location to play next. Each two-hop transmission made by the relay on location is naturally described by a Bernoulli trial defined as:

$$X_i = \begin{cases} 1 & \text{if transmission success;} \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

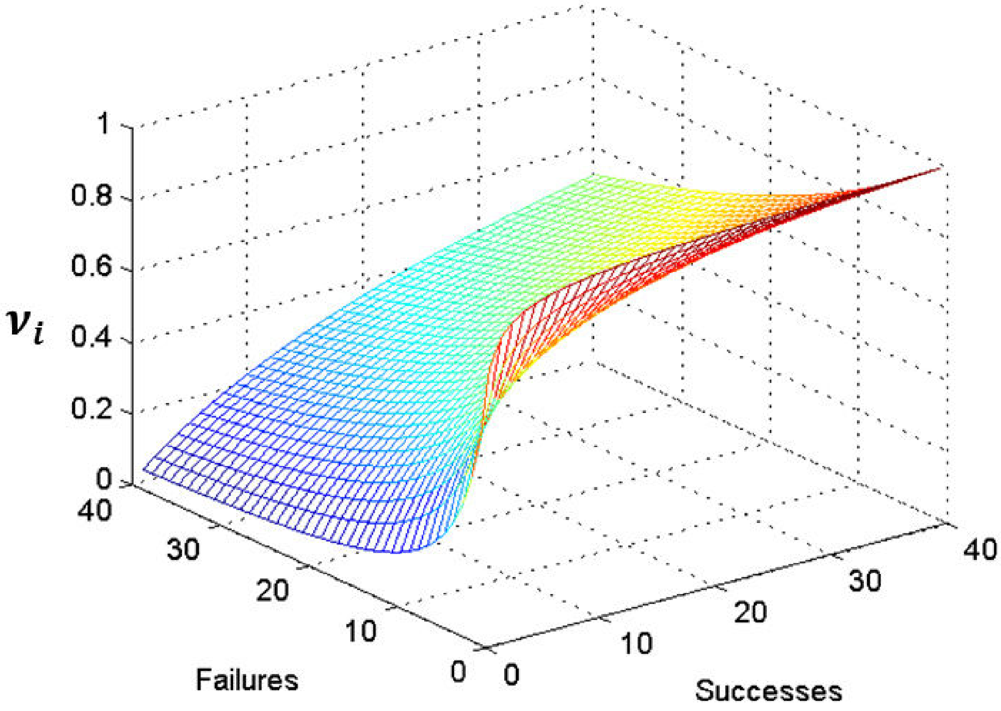


Figure 2-2: Gittins indices computed for a Bernoulli reward process (N=750, β=0.95)

A computational method for calculating indices for a Bernoulli reward process is described in Gittins [25]. Briefly, the infinite horizon is approximated with a large finite horizon, and backwards induction is used to solve for indices. The state vector in this case comprises simply n_i, the number of plays on this location, and s_i, the number of successes at this location: the index is thus $\nu_i(n_i, s_i)$, which can be stored as a lookup table. The indices are computed in real time by the relay and updated as new information becomes available. Fig. 2-2 shows the Gittins indices computed for the Bernoulli process, the index decreases logarithmically as a function of the total observations as well as the number of failures such that unknown arms are prioritized first, but the performance of each arm becomes increasingly important.

For the MAB autonomous experiment, we heuristically account for switching costs by designating five transmissions as one observation, so that the time spent on location is at least as long as the shortest transit time away. We define the reward $\bar{\theta}_i$ of each machine as the estimated mean of the Bernoulli random variable $X_i(t)$ for those five transmissions. The estimate of the mean and variance of $X_i(t)$ can be re-computed with each new sample as a function of n_i, following [39]:

$$\bar{\theta}_i(t) = \begin{cases} \dfrac{n_i - 1}{n_i}\bar{\theta}_i(t-1) + \dfrac{1}{n_i}X_i(t) & \text{if } i \text{ was played} \\ \bar{\theta}_i(t-1) & \text{otherwise.} \end{cases} \tag{2.10}$$

$$\hat{\sigma}_i^2(t) = \begin{cases} \dfrac{n_i - 2}{n_i - 1}\hat{\sigma}_i(t-1) + \dfrac{1}{n_i}(X_i(t) - \bar{\theta}_i(t-1))^2 & \\ \qquad\qquad \text{if } i \text{ was played and } n_i \geq 2 & \\ \hat{\sigma}_i^2(t-1) \qquad\qquad \text{if } i \text{ was not played.} & \end{cases} \tag{2.11}$$

Thus, $\bar{\theta}_i$ is the best estimate of the probability of packet success at location i — a practical measure of the acoustic channel's performance that can be updated with incoming samples. Its standard deviation $\bar{\sigma}_i$ is given by:

$$\bar{\sigma}_i^2 = \frac{1}{n_i}\hat{\sigma}_i^2, \tag{2.12}$$

which determines the weighted benefits of exploration as defined by the index calculation. By the Central Limit Theorem, assuming the Bernoulli trials are independent, the reward distribution approximates a normal distribution as the number of observations gets large. Gittins [24] showed that the index for this reward process is a function of the mean (expected reward) and its standard deviation (uncertainty of estimate). This is expressed as:

$$\nu(\bar{\theta}_i, n, \bar{\sigma}_i) = \bar{\theta}_i + \bar{\sigma}_i \nu(0, n, 1). \tag{2.13}$$

$\nu(0, n, 1)$ are Gittins indices for the normal distribution with zero mean and unit variance, and have been previously tabulated [25]. Indices are stored in a lookup table and accessed in real time by the relay, which updates as new information becomes available. Fig. 2-3 shows Gittins indices $\nu(0, n, 1)$, which decrease logarithmically with the number of observations.

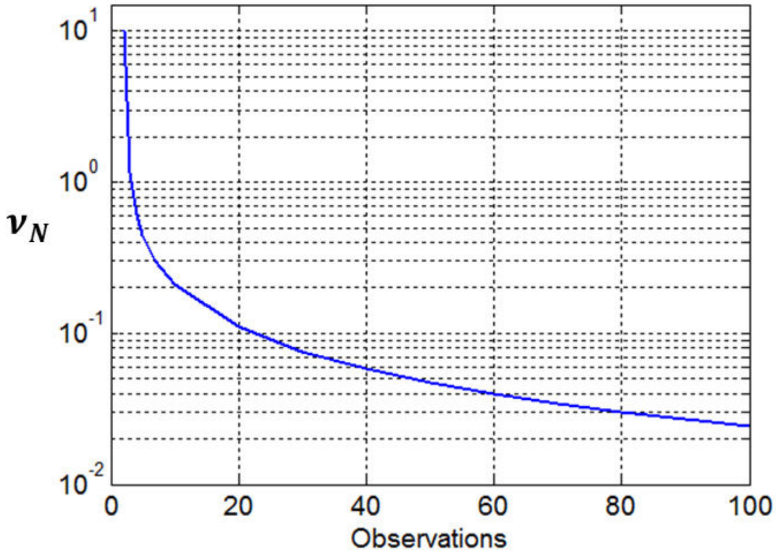


Figure 2-3: Gittins indices for Normal reward distribution with zero mean and unit variance.

Chapter 3

Field Implementation and Experiments

We consider a one-way, two-link acoustic transmission in the Charles River Basin, Boston MA. A source modem located at the MIT Sailing Pavilion broadcasts a data message, which is repeated by the relay; an acoustic modem towed by a robotic surface vehicle. The destination node is a second robotic vehicle station-keeping 580m across the river from the source. A transmission is considered successful only if both hops succeed, i.e., the relay decodes the source packet, and the destination decodes the relay packet. Source transmissions may reach the destination directly; this through-transmission success rate reflects the performance of the acoustic channel with no relay. In setting up the adaptive positioning experiments, we assumed no prior knowledge of the acoustic channel beyond the usual spreading law. Nine candidate relay locations were chosen in a grid pattern centered on the line between the source and destination nodes (Fig. 3-1). In practice, such a choice would be influenced by mission constraints. For all experiments described, Site 1 was designated the starting location.

We describe two types of experiments, "autonomy" trials and "hybrid" trials. Autonomy trials were conducted with the Gittins index MAB algorithm and the adapted MABSC algorithm implemented as fully autonomous, turn-key elements in an autonomous multi-vehicle systems. Both acoustic acknowledgments, required by

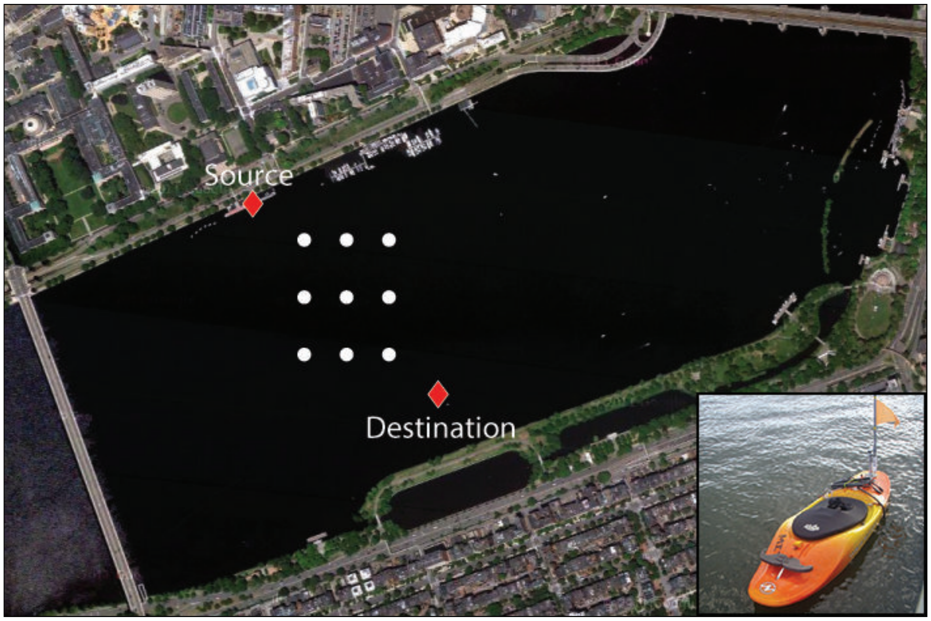


Figure 3-1: Charles River Basin (Boston, MA) with Autonomous Surface Vehicle *Nostromo* inset. Relay locations are shown in white. Source and destination locations are shown in red.

fully underwater systems, and WiFi acknowledgments (for simplicity) were implemented. Hybrid trials were touring surveys that transmitted a set number of times at each point in turn, for the purposes of building a large dataset from which datapoints can be sampled offline.

3.1 Autonomous Surface Vehicles

All field experiments were conducted with custom-built autonomous surface vehicles as the relay and destination nodes (Fig. 3-2). The source was a fixed station at the MIT Sailing Pavilion with a modem at the dock. Our autonomous surface vehicles tow acoustic modem transducers at a fixed depth to simulate underwater communications, with the benefits of GPS and WiFi connectivity for controlled experiments. Here, we describe the main components of the autonomous system, and further hardware and software details are presented in Appendix A.

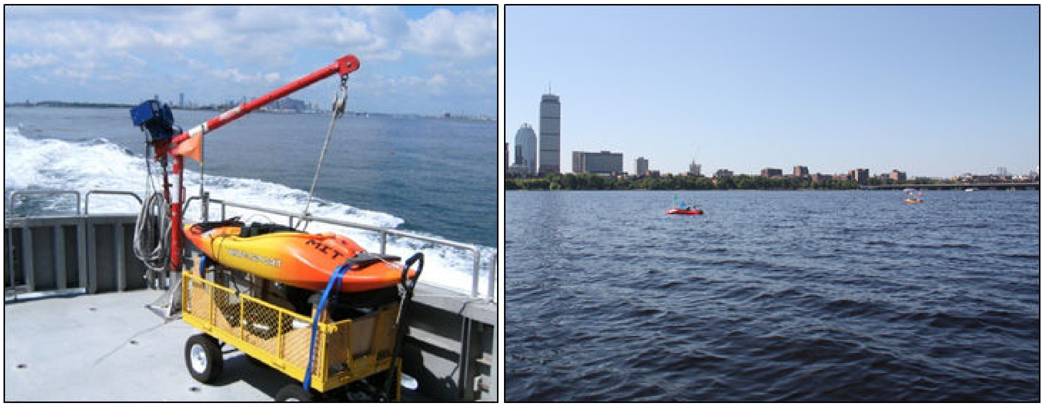


Figure 3-2: Autonomous surface vehicle *Nostromo* (left) in Boston Harbor and with *Silvana* (right) on the Charles River.

3.1.1 Vehicle Hardware

The vehicles, built on small whitewater kayaks, are 1.8m (5.9ft) in length and weigh roughly 40kg (88lbs). A bow-mounted trolling motor provides 220N (55lbs) of thrust for a maximum speed of 3m/s (6 kts), making the kayaks easy to control and highly maneuverable. For the purposes of the experiments, the vehicles are commanded to travel a constant 1.5m/s and maintain a station-keeping circle ten meters in diameter on location. Two vehicles were used in field trials, as well as a fixed shore station *Icarus* located at the MIT Sailing Pavilion.

3.1.2 MOOS-Based Software Architecture

MOOS (Mission Oriented Operating Suite) and MOOS-IvP software [8] are robotics middleware packages developed by Paul Newman and Mike Benjamin to assist software development on robotic platforms. MOOS provides a basic message-passing and database service that allows multiple programs running on different platforms to share information in an organized manner, while MOOS-IvP adds applications specifically designed for the needs of marine vehicles. In particular, each vehicle maintains a personal communications database called the MOOSDB. A MOOS process (MOOSApp) can publish data to this database, or subscribe for an update each time a variable is published, using asynchronous thread-based operations. In this way, communication between processes do not have to be handled individually.

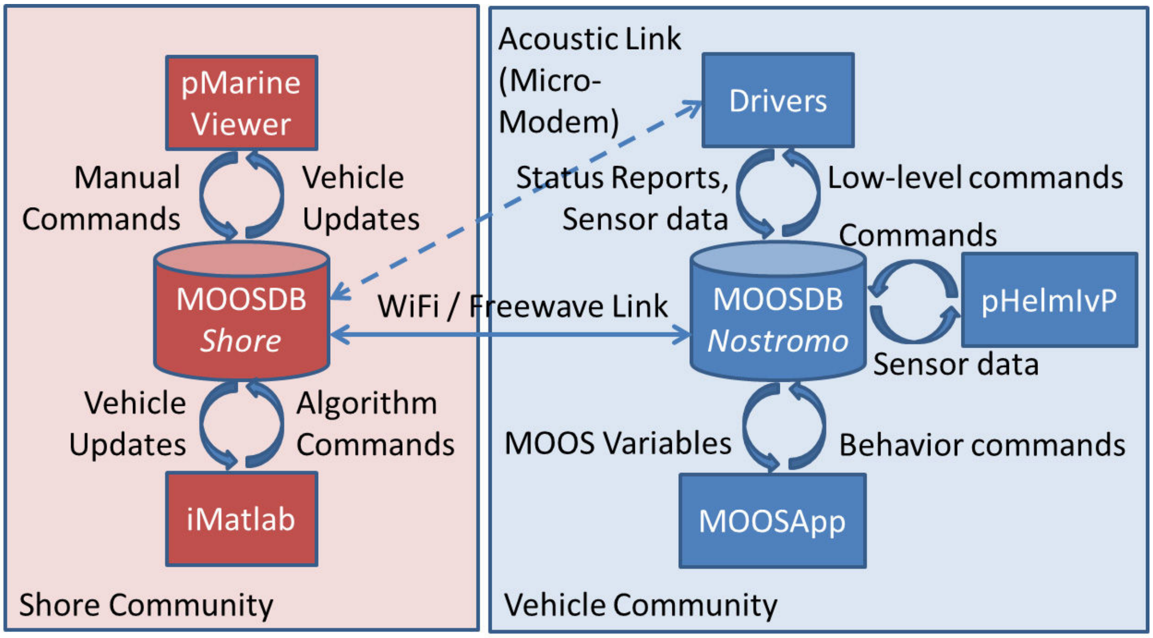


Figure 3-3: Illustration of MOOS and MOOS-IvP software architecture used for communications and control.

MOOSIvP's *pHelmIvP* is used as a "back-seat driver" on the vehicles, making high-level autonomy decisions based on pre-determined behaviors and communicating with low-level vehicle drivers. Multi-objective optimization is used to select a target course, speed and depth (if applicable) if there are competing behaviors active. For this application, we mainly implement the waypoint behavior and the station-keeping behavior. For the waypoint behavior, an x,y co-ordinate is commanded and a trackline is generated between its current and desired locations. In order to reduce crosstrack error, MOOS-IvP's trackline behavior then modulates the vehicle's desired heading so as to steer it towards a point on the trackline some lead distance ahead. The closer the vehicle is to the trackline, the further the lead distance is along the trackline. Station-keeping behavior turns off the vehicle's motor when it is within some inner radius, and allows the vehicle to drift beyond some outer slip radius before restarting the thruster. For all field experiments considered, the inner radius was three meters and the slip radius was ten meters. We implement open-source software library RTKLib (Real-Time Kinematic Library) [48] in conjunction with a GPS base station at the MIT Sailing Pavilion to achieve GPS noise covariance on the order of 10cm^2.

Onboard MOOSApps were implemented for the purpose of autonomy experiments. Gittins indices were computed offline and stored in a look-up table provided to the relay. Policy indices were computed on the shoreside computer with Matlab and communicated to the vehicles for simplicity; the computation itself could be handled in C++ with a slightly more powerful processor than the current single-core Gumstix onboard.

3.1.3 Acoustic Modems

We use Woods Hole Oceanographic Institution (WHOI) Micro-modems [22], an established and commercially available technology for underwater acoustic data transmission (Fig. 3-4). The Micro-modem transmits at a fixed frequency (25kHz) and power (50W burst for variable duration dependent on packet type). NMEA 0183 messaging is used for commands and communication with the Micro-modem, and several transmission rates characterized by different error correction codes and modulation types are available.

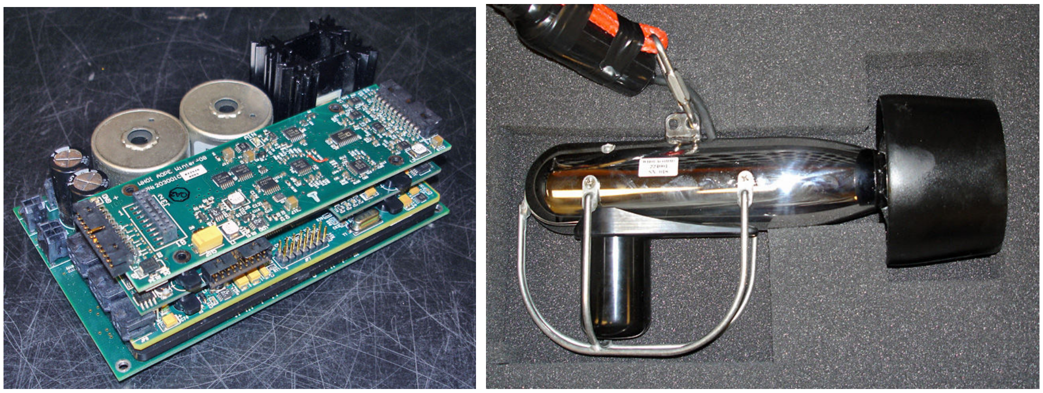


Figure 3-4: Micro-modem Multi-Channel PSK Stack with Power Amplifier (left) and 25kHz Omnidirectional Transducer Towfish (right) Source: `http://acomms.whoi.edu/umodem/`

For the purpose of field experiments, quadrature phase-shift-keying (QPSK) rates 1 and 2 were used mainly for the reasonable tradeoff between packet success rate (in the range of 40% to 97%), packet data payload and time taken to transmit. Higher rates are characterized by larger packet payloads and correspondingly longer transmission times and lower packet success rates (See Table B.2 in Appendix B). We note that programmable modem parameters such as packet encoding scheme can be included combinatorially as additional machines in the multi-armed bandit, but we have fixed these for simplicity. The Micro-modem reports various transmission statistics, and we present signal-to-noise ratios (SNR) from before the equalizer on the receiving modem ("SNR-In") as a representation of the physical channel quality. Statistics of other available measures (e.g. SNR out of the equalizer) are presented in Appendix A. A two-hop transmission as described above takes a minimum of fifteen seconds with this hardware.

3.2 Autonomous MAB

For the MAB algorithm, a sample was designated as five transmissions so that the time required to sample once was comparable to the time of transit to another location. All data transmissions were sent at phase-shift keying (PSK) Rate 2, with a fixed message size of 192 bytes. Estimates of the packet success mean and its variance were initialized with a touring survey consisting of two samples at each location. This initialization was required as at least two observations are required to calculate a Normal Gittins index. Subsequently, the relay executed the algorithm, selecting the location with the highest index and breaking ties by favoring shorter travel distances. Two experimental trials were conducted on the same day. Acknowledgments of transmission receipts were communicated from the destination to the relay over WiFi for simplicity.

3.3 Autonomous MABSC

We employed the Bernoulli reward scheme with each sample consisting of a single two-hop transmission sent at PSK Rate 1. The MABSC algorithm was initialized with the assumption of 100% packet success probability at each site[1]. The look-

[1] Practically, the choice of initialization represents an acceptable performance threshold. Unexplored sites may never be chosen if a previous site maintains performance above or equal to the

ahead horizon for policy enumeration was constrained to a maximum computation time of fifteen seconds. For a look-ahead horizon of five, the average computation time is on the order of one second[2]. For an underwater vehicle, learning the result of the relayed transmission (success or failure) from the destination robot must be done with acoustic acknowledgments. To simulate this with our surface vehicles, we utilized the Micro-Modem frequency-shift-keying (FSK) Mini-packet, a 13-bit message with robust performance. Our experiments have consistently shown packet loss rates of less than 5% for the FSK Mini-packet, thus it is substantially more reliable than the PSK 192-byte packets used for data transmission. If the acknowledgment is lost, the two-hop transmission is considered a failure by the relay. A touring survey consisting of a single circuit with ten transmissions at each site was performed before the experiment to provide a comparison measure, and the MABSC algorithm was run for the same mission time (55 minutes). Subsequently, the relay robot executed the MABSC algorithm autonomously.

3.4 Hybrid Dataset Touring Surveys

In the field it is difficult to compare the performance of several competing algorithms as multiple relays would share the same physical space and channel, resulting in transmissions experiencing acoustic interference or extended wait times. Conducting experiments on different days is also undesirable as changing weather and surface conditions make it difficult to objectively evaluate the improvement in performance due to action by the algorithms. Thus, we construct a hybrid experiment; first, by collecting a large dataset of transmissions on a single experimental day. A touring survey taking five transmissions at every location was conducted for several hours. Then, each decision algorithm was applied to the same dataset, i.e. transmission results were sampled from the dataset for the appropriate time and location and used to update the algorithm's information state. The shallow-water acoustic environment

threshold. Here we have prioritized exploration of all possible locations.

[2] Computed with Matlab R2012b on Windows 7 (64bit), Intel i5-3450, 16GB of RAM

is in general difficult to model and using field data allows us to capture complex spatially-dependent behavior. The hybrid dataset contained 835 detected transmissions from source to relay and 636 detected transmissions from relay to destination, with 493 of these being successfully decoded relayed transmissions.

Chapter 4

Acoustic Communications in Shallow Water

The transmission power of the Micro-modems is capable of transmitting many kilometers in the open ocean, however, in a shallow-water man-made environment like the Charles River, the performance of the channel is limited by multipath interference. Fig. 4-1 shows altimetry data revealing irregular bottom topography in the area visited by the relay, especially a shallower shelf to the northeast. Though not visible, a deeper channel is also present towards the south (Boston) bank where the destination node is situated. Furthermore, there is a long stone wall about 10m behind the source node and a hard seawall on the opposite bank. The depth of the water in the Charles River basin is controlled by a series of locks at its mouth in Boston harbor. The MAB formulation is set up to adaptively explore these space-varying properties of the acoustic channel without sacrificing overall data transmission or requiring prior knowledge of the channel characteristics.

Fig. 4-3 shows SNR-In values reported for all acoustic transmissions during the MAB field experiment, Trial 2. The spread of values is -wide (25 to 25dB) and there is no clear spatial structure to the distribution. The closest locations for each respective link (Sites 1 and 5) do not have discernibly higher SNR-In values. Despite relatively high SNR-In values, the complex, shallow-water environment makes packet decoding difficult and introduces the spatial variation that the multi-armed bandit exploits.

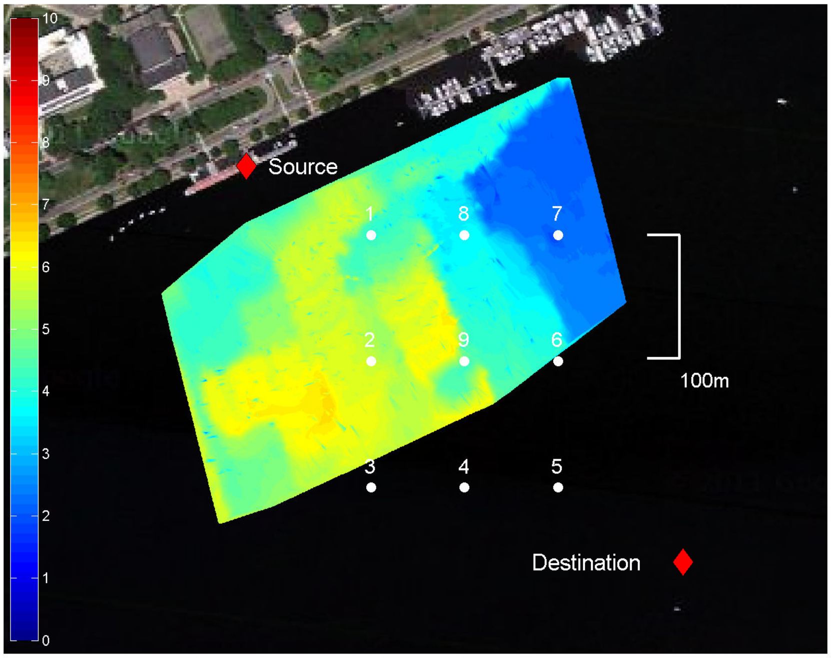


Figure 4-1: Altimetry data of the Charles River, Boston in the area visited by the relay. Scale shown is from 0 to 10m. The depth of the Charles River varies from 2 to 12m, with the deepest area in a channel on the South (Boston) side of the river.

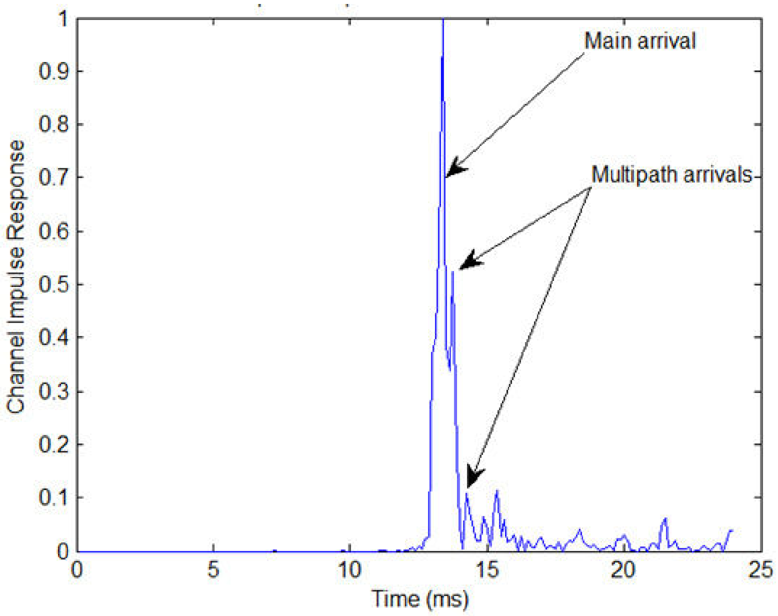


Figure 4-2: Impulse response of the matched filter showing various multipath arrivals after the main signal. The mean squared error of the modem's equalizer is another measure of the interference in the channel (Appendix B).

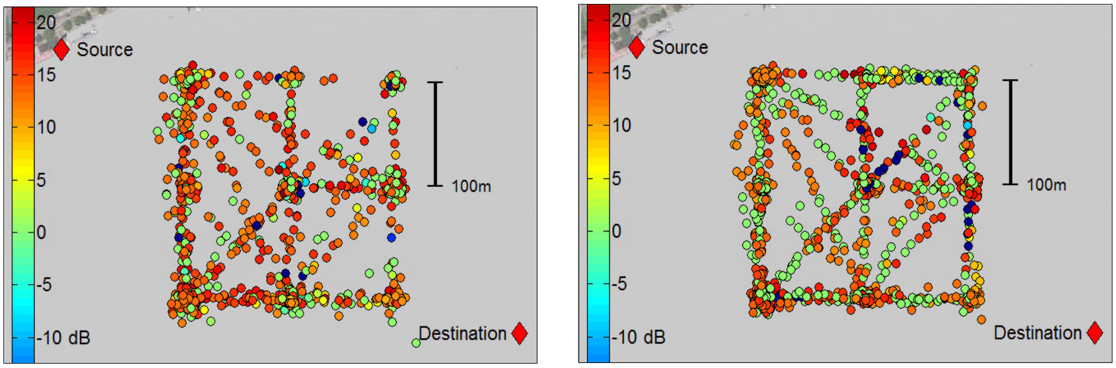


Figure 4-3: SNR-In for data transmissions from MAB field experiment, Trial 2. Direction of transmission is shore to relay (left) and relay to destination (right).

The SNR of the initialization touring survey is shown in Fig. 4-4 for each link of the two-hop transmission. The range shown is from 0 to 20dB, and high variability in SNR values as well as the horizontal space is observable. When compared with the final packet success rates (Fig. 4-5), it is not clear that the SNR, a traditional measure of channel performance, is accurately correlated with the packet success rate on location.

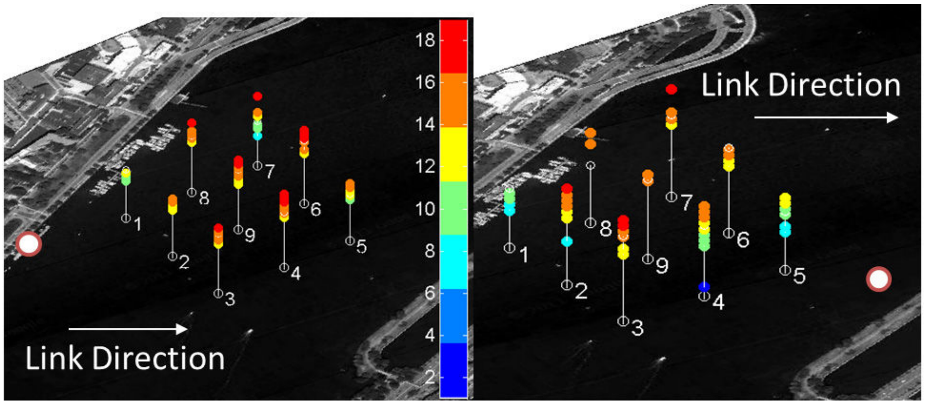


Figure 4-4: SNR-In for initial ten transmissions at each of nine relay locations. Range is from 0 to 20dB.

Fig. 4-6 shows SNR-In values and final packet success rates from the hybrid experiments (HybridSetA and B) of the nine relay waypoints conducted on different days. For these surveys, the mobile relay transmitted five times at each location per visit, at PSK Rate 2 with a fixed message size of 192 bytes, and acknowledgments of transmission receipts were communicated over WiFi for simplicity and shorter cycle

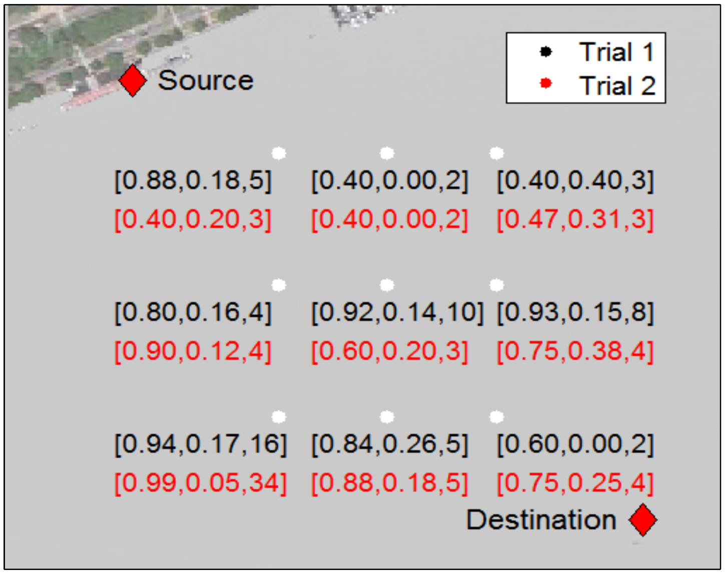


Figure 4-5: Final [packet success rates, variance of the estimate, number of samples] at nine bandit locations and for both autonomy trials.

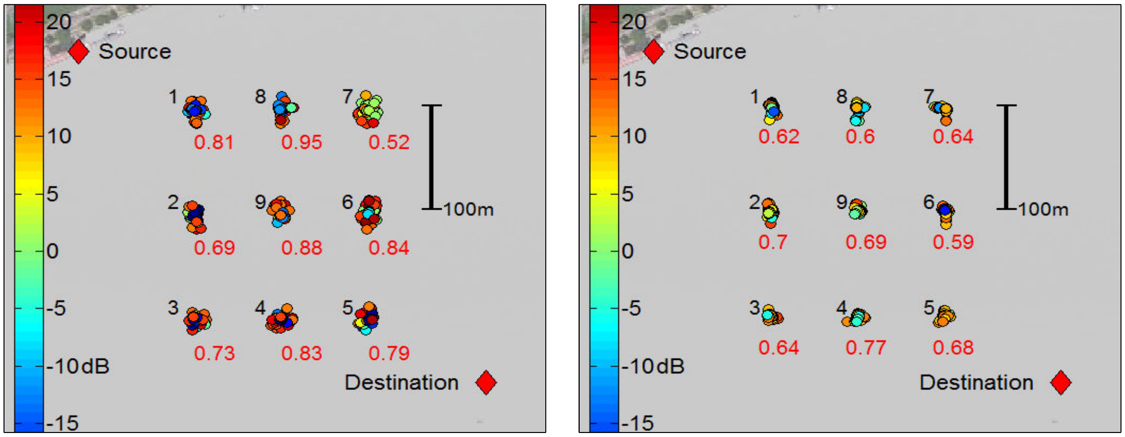


Figure 4-6: SNR-In for data transmissions from HybridSetA (left) and B (right). Direction of transmission is source to relay. Site number is shown in black and final packet success rates estimates over the whole mission is shown in red.

times. There is significant spatial variation in final packet success rates by location. HybridSetA showed a spread in means of 43%, with Site 8 outperforming the next nearest by 7%. In comparison, HybridSetB shows a narrower spread of 18% and Site 4 was the highest performing and Site 8 exhibiting poor performance. Variation from day to day motivates the need for a fast and adaptive learning algorithm without the need for re-tuning heuristic parameters. The MAB and MABSC formulations adaptively explore these space-varying properties of the acoustic channel without sacrificing overall data transmission or requiring prior knowledge of channel characteristics. The SNR-In values for the source-to-relay transmission link from each dataset is presented in Fig. 4-7 (the same figures for the relay-to-end link is presented in Appendix A). No clear trend can be distinguished either in time or space, and the variation is approximately -20 to 20dB.

SNR-In values for the five transmissions taken each hybrid experiment were averaged and Fig. 4-8 (left) shows the progression in each site for the time-averaged values of SNR-In. There is no clear trend in these values temporally and thus we assume the Bernoulli transmission processes to be acceptably stationary over the time scale of the experiment. Remarkably, as illustrated in Fig. 4-8 (right), there is essentially no correlation of SNR-In with the corresponding grouped packet success rates of those transmissions, with high variation in SNR-In even for 100% success. These statistics of the acoustic channel from our field experiments support the assumption of stationarity for the multi-armed bandit, i.e., the channel characteristics do not change meaningfully over the course of the mission and the Bernoulli processes can be assumed stationary. The mean squared error (MSE) of the equalizer as well as the SNR-Out of the equalizer are additional statistics given by the modem that measure the interference of the channel and the performance of channel estimation and error correction schemes. These statistics are presented in Appendix A.

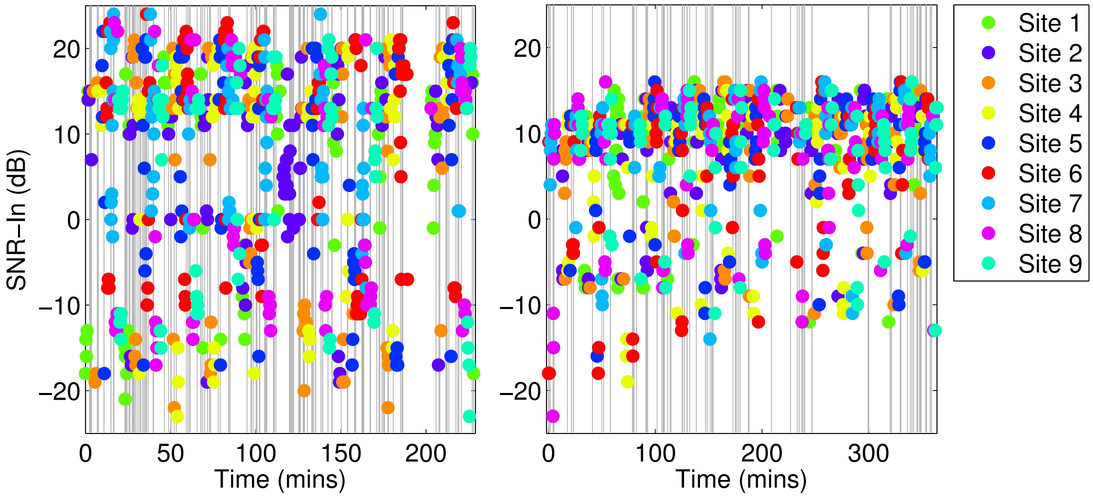


Figure 4-7: SNR-In values over time for HybridSetA (left) and B (right). Grey lines represent lost packets. Sites are visited in the same order and data is for source-to-relay transmission.

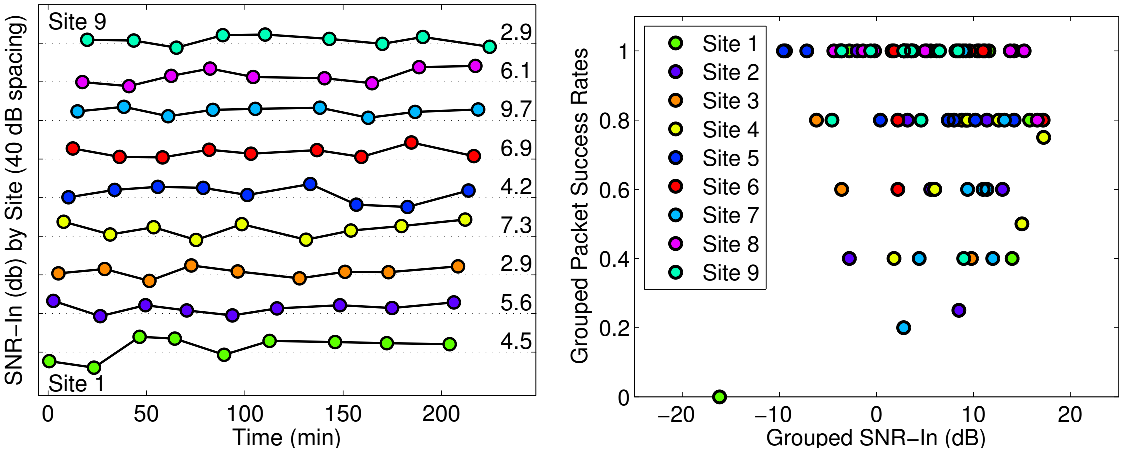


Figure 4-8: HybridSetA Grouped SNR-In values over time (left), with overall average noted at right, and Grouped Packet Success Rates against SNR-In (right). Data is for source to relay transmission only. Each averaging group consists of five transmissions on location.

Chapter 5

Experimental Results

We evaluate each algorithm's performance in achieving the overall goal of maximizing the cumulative bytes of data successfully relayed. While all of the algorithms considered except for ϵ-greedy are asymptotically efficient and converge to the optimal point given enough observations, the short-term performance of each algorithm is of significant practical importance.

5.1 Autonomous MAB

As noted in Chapter 3, each sample corresponds to five transmissions that are Bernoulli trials. For comparison with a method that does not take into account location-dependent acoustic performance, we extrapolate the expected reward from the initial touring survey and present the difference in reward obtained, in Figs. 5 and 6. During the exploration regime following the initialization period, MAB decision-making performs similarly or worse than the touring approach as the vehicle continues to visit locations with poor performance but higher standard deviations.

Fig. 5-1 presents the evolution by observation of the algorithm's three key parameters for Trial 2: the Gittins index, the estimated mean, and the variance of the estimate of each machine. The initial choice of exploration for Location 7 with low mean but high variance is clear, as well as eventual settling into the high-performing Location 3. As the number of observations increases, the Normal Gittins index falls

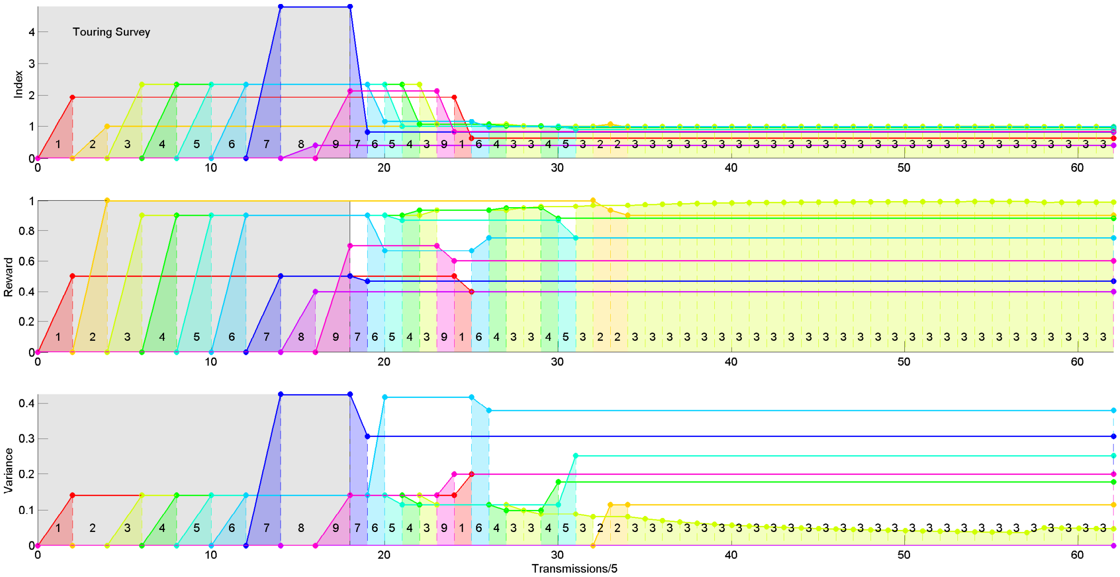


Figure 5-1: Evolution of the Gittins Index, the mean, and the variance of the estimate for Trial 2.

away exponentially, and the algorithm asymptotically favors one or several points with comparatively higher performance. Evaluation of the algorithm's performance is considered in terms of cumulative bytes of data successfully transmitted via the relay (Fig. 5-2). For comparison with a method that does not take into account location-dependent acoustic performance, we extrapolate the expected reward from the initial touring survey and present the difference in reward obtained, in Fig 5-3.

During the exploration regime following the initialization period, MAB decision-making performs similarly or worse than the touring approach as the vehicle continues to visit locations with poor performance but higher standard deviations. For Trial 1, the cumulative data transmitted eventually exceeds that of the touring extrapolation by 14% and the final data transmission rate is improved by 33%, while for Trial 2, the cumulative data transmitted and the rate of transmission increased by 19.6% and 37% respectively. In both trials, the MAB behavior was effective at improving cumulative data transmission as compared to the initial survey. However, approximately 60% of mission time was spent in transit, highlighting the significance of switching costs.

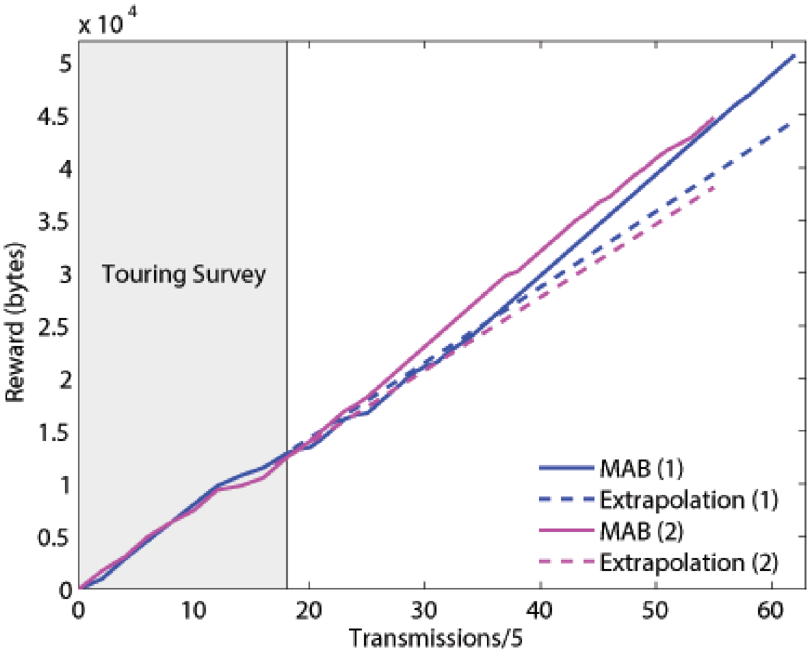


Figure 5-2: Cumulative bytes transmitted for Trials 1 and 2 and extrapolations of initial touring survey.

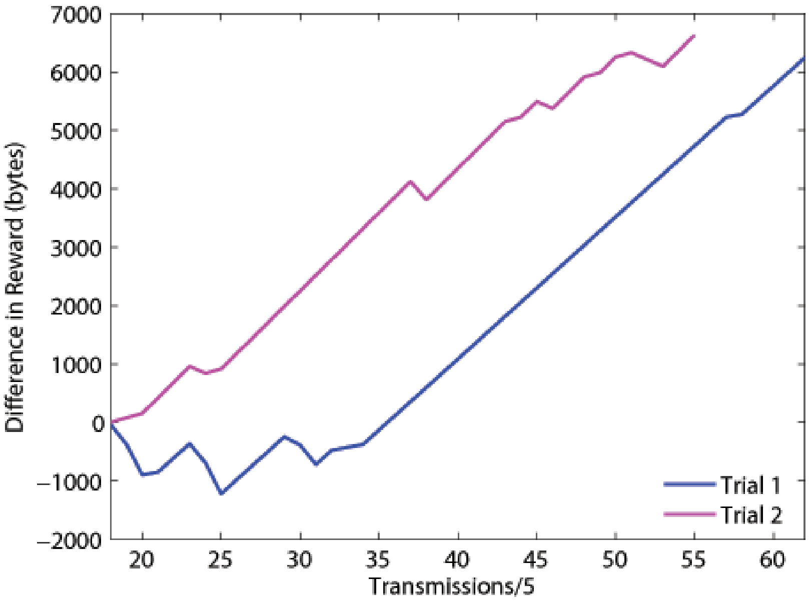


Figure 5-3: Improvement in cumulative bytes transmitted with respect to extrapolations of initial touring survey.

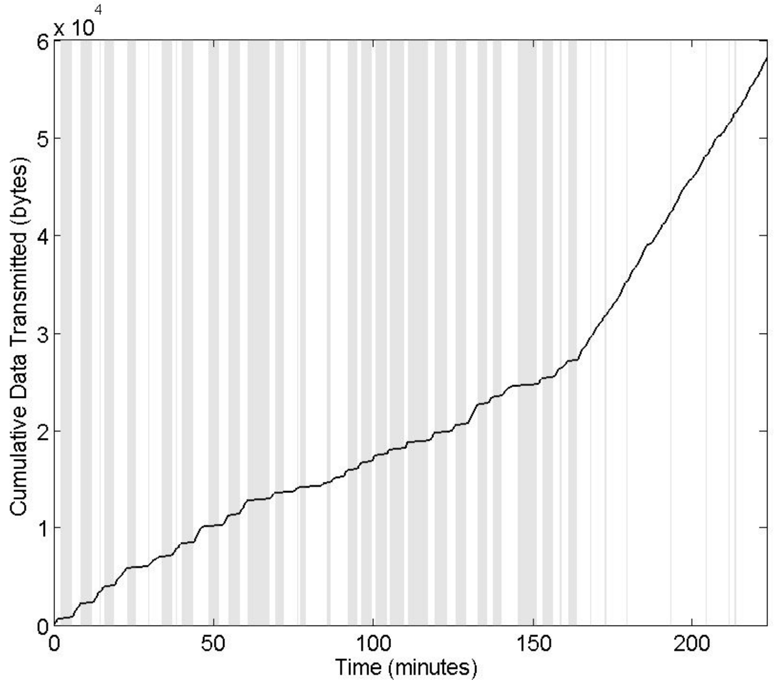


Figure 5-4: Cumulative data transmitted as a function of time in minutes for Trial 2, with periods of vehicle transit shown in gray.

5.2 Autonomous MABSC

We compare the performance of the autonomous MABSC decision algorithm to the touring survey conducted on the same day and consisting of ten transmissions at each site over one circuit. The two algorithms were limited to the same mission duration. Figs. 5-5 and 5-6 show the performance of the MABSC in comparison with the tour. During the first thirty minutes of the mission, performance was similar as the MABSC explored the unknown field. For the remainder of the mission, the MABSC settled to a high-performing site, and achieved a final reward rate 77.2% higher, and an average reward rate 28.4% higher than the touring survey. Since the MABSC is able to discard poorly-performing locations while accounting for the cost of switching locations, it was able to carry out more transmissions (Fig. 5-5).

We have shown that the MABSC decision policy efficiently maximizes data transmission and outperforms a simple touring survey. Finally, we recall that the multi-armed bandit formulation optimally trades-off exploration and exploitation, and we

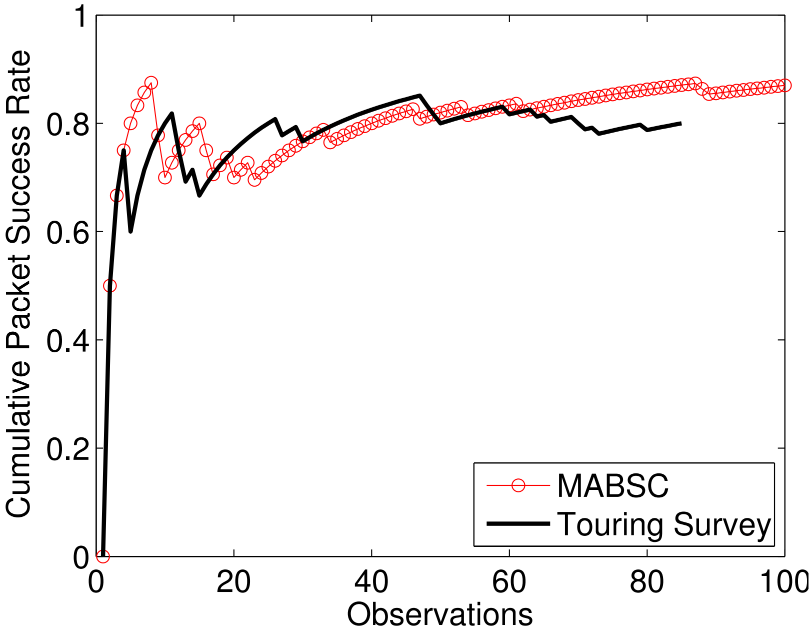


Figure 5-5: Cumulative packet success rate by observation, where unity indicates 100% success rate.

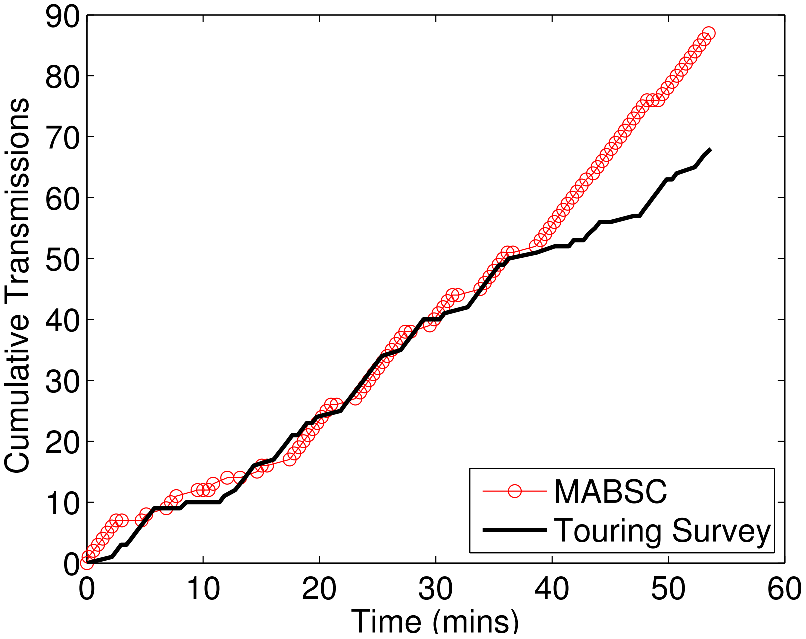


Figure 5-6: Cumulative data transmitted by mission time.

expect the MABSC to gain information about the field effectively. Thus, we assess the performance of the touring survey and the MABSC policy in exploration. We value an algorithm's information gain by computing the total sum of squares error for each algorithm summed over all relay sites. The squared error is taken between the algorithm's current estimate of the site's performance, and the best possible estimate obtained from both datasets. Fig. 5-7 shows the evolution of this error by observation for the touring survey and the MABSC. It is clear the MABSC gathers information about the performance field significantly faster than a simple touring survey, and does so without sacrificing data transmission.

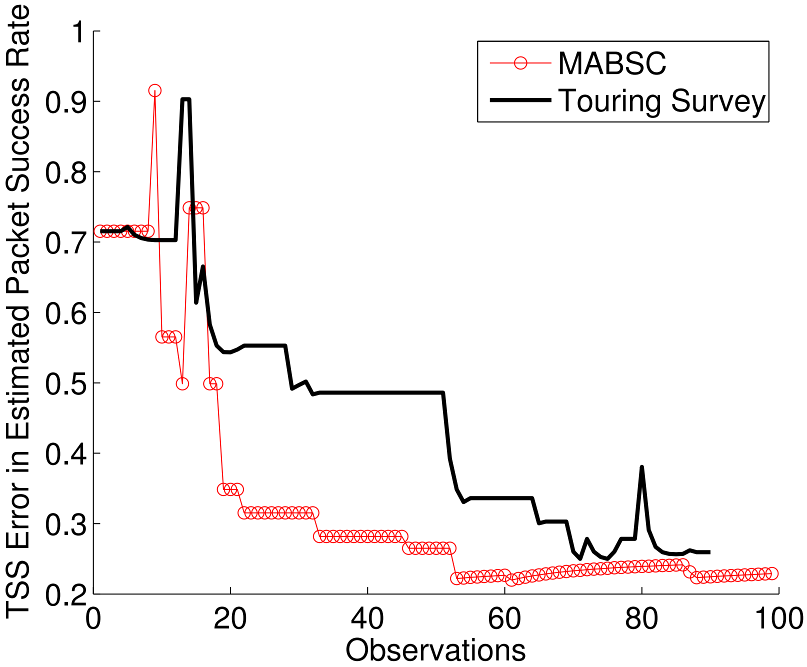


Figure 5-7: Total sum of squares error by observation.

5.3 MABSC on the Hybrid Tour Data

We now compare throughput performance of the MAB, the switching cost adaptation for MABSC, ϵ-greedy and ϵ-decreasing algorithms, and the touring survey. An ϵ-greedy algorithm plays the best arm $(1 - \epsilon)$ of the time and switches to a random arm ϵ of the time. ϵ-decreasing is a variation on ϵ-greedy where the value of

ϵ decreases in time. Algorithms other than the MABSC have not been adapted to account for switching costs, and there was no restriction on the number of switches for any algorithm. ϵ-dependent algorithms were tuned with ϵ and τ of differing orders of magnitude; only a high-performing subset is presented here for the sake of brevity. Transmissions were sampled from the tour dataset in chronological order, terminating when unavailable data was requested. Since the number of transmissions at each location is limited by the total mission time of the touring survey conducted in the field, fewer observations are generated for greedier algorithms that sample at fewer locations, i.e., perform less exploration. We evaluate each algorithm's performance in terms of the average packet success rate achieved, attained from the cumulative number of successful transmissions.

Fig. 5-8 shows the performance of each algorithm as a function of observations (transmissions). The estimated best-site and average success rates were computed from each dataset as a whole.

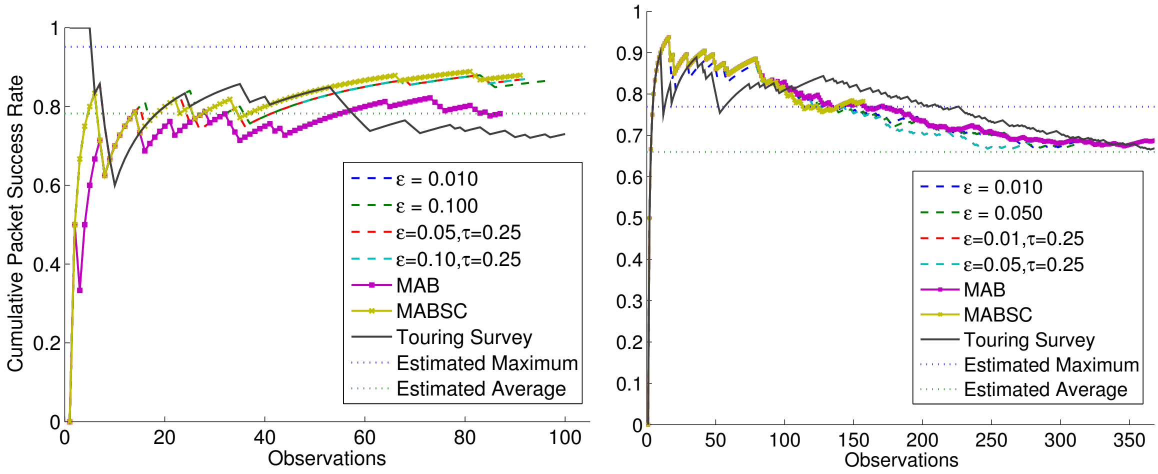


Figure 5-8: HybridSetA (left) and B (right) cumulative performance of MAB, MABSC and tuned ϵ-greedy and ϵ-decreasing algorithms by observations, where unity indicates 100% success rate.

In HybridSetA, as before, the canonical MAB shows poorer relative performance at least in the early stages, by first learning about all sites. In comparison, MABSC actively avoids switching, and is closely competitive with tuned ϵ-greedy and ϵ-decreasing algorithms. The overall trend was an increase in performance with the number of observations. On HybridSetB, the algorithms were similar in performance. The touring survey initially performed worse than others, becoming very good around 100 observations, and deteriorating again towards the end. This behavior can be attributed to high initial estimates of channel performance, with the exploitative algorithms sampling more from high-performing locations and driving down their cumulative performance at a quicker rate, while the touring survey continues to sample all locations evenly. Although performance of the MABSC and ϵ-greedy methods are comparable on a per-observation basis, the impact of switching times is substantial; see Fig. 5-9, which accounts for transmission and transit times.

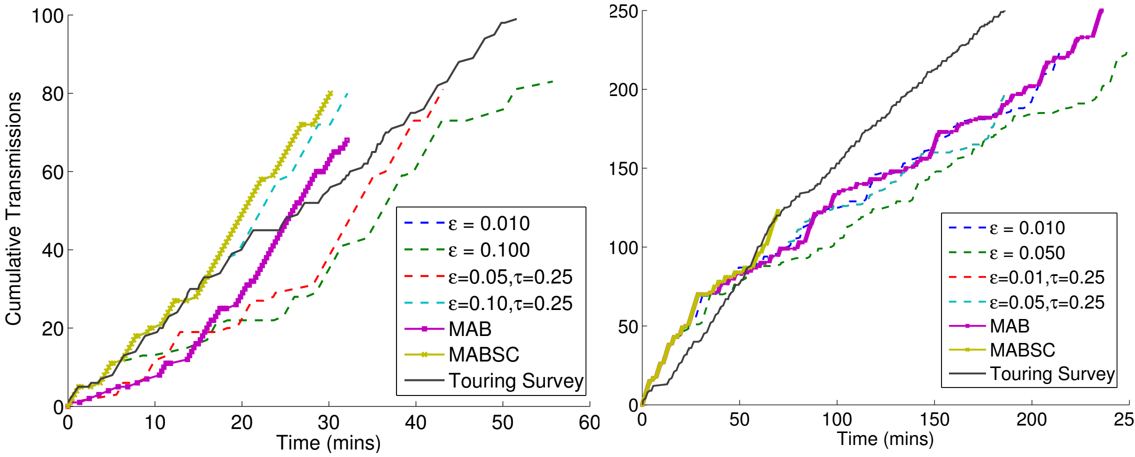


Figure 5-9: HybridSetA (left) and B (right) cumulative transmissions by calculated mission times. In left, the plot of $\epsilon = 0.010$ overlaps with $\epsilon = 0.1, \tau = 0.25$.

In HybridSetA, with high performance on the best site, ϵ-greedy with the greatest value of ϵ demonstrates a slow overall rate as expected, while decreasing values of ϵ demonstrate higher rates. The direct MAB formulation is competitive in real time and its performance is significantly improved by the adaptation to switching costs. In HybridSetB, with narrow performance ranges, the algorithms which do not take into account traveling time perform worse in the short-term than the touring survey (constant five transmissions per site). All except for $\epsilon = 0.05$ eventually settle on the highest performing point. Noticeably, the MABSC algorithm terminates early, around 75 minutes. It has very quickly settled on the highest-performing location and taken all possible samples in that location from the dataset.

The MABSC algorithm performed enumeration for 17% of decisions, made use of Asawa's theorem to continue without computation 37% of the time, and found, with trivial computation, that the current point had the highest Gittins index 46% of the time. At a computation horizon of size six, enumeration takes 1.95% of total mission time, as compared to 7.8% without using the switching index and 11.3% if enumerating at every decision point.

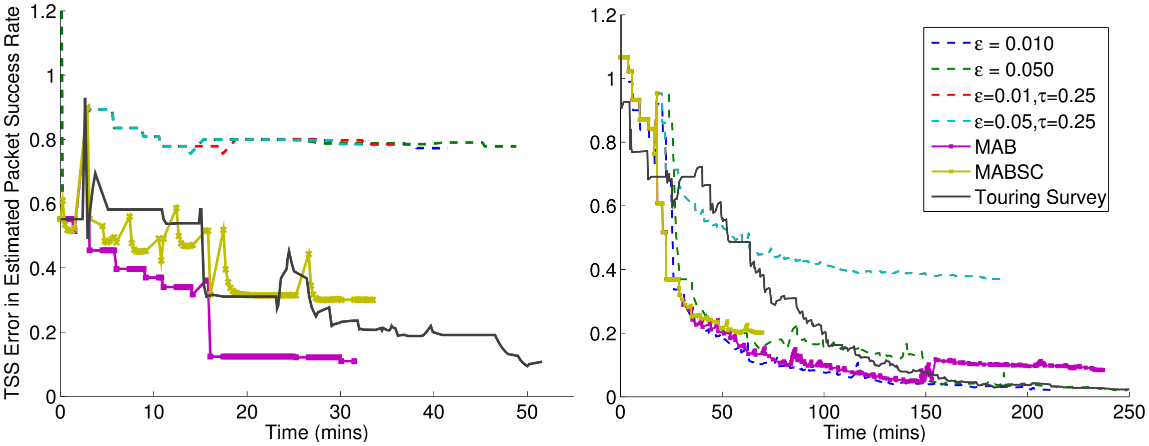


Figure 5-10: HybridSetA and B Total Sum of Squared Differences over time, summed over all sites.

As expected, the MAB improves the estimate at each site most efficiently, arriving at an extremely good estimate in a low number of observations. In contrast, the ϵ-greedy and ϵ-decreasing algorithms give very poor estimates of the whole field. The MABSC provides a good compromise between information-gathering and maximizing reward.

5.4 MABSC with Synthetic Data

Using synthetic data allows us to investigate the asymptotic performance of these algorithms with different probability spreads, enumeration horizons and switching cost models, however, the short-term behavior demonstrated in field experiments is more relevant to practical applications in ocean systems. Packet success probabilities were randomly assigned to the nine sites in two structures: a narrow range equally spaced between 0.7 and 0.8, and a wide range equally spaced between 0.1 and 0.9.

We average the results of 100 trials, each trial consisting of 450 observations for each algorithm. Bernoulli data was generated to match site probabilities and randomly permuted. Table 5.1 shows the average percentage improvement in reward rate (data transmitted per unit time) over a touring survey, for the MAB and MABSC with several lookahead horizons. The touring survey takes five samples at each location, for a total of ten circuits. We use rate of reward for comparison as it factors in the time cost of switching location. The MAB improves significantly over a simple ϵ-greedy strategy or touring survey in both cases, and the MABSC provides further gains although with apparently diminishing returns. The computation time required for a single decision enumeration is reported where applicable[1]. These results with longer enumeration horizons are very likely to be better than those which would be provided by the one-step lookahead MPI or LP restless bandit solutions.

We also consider three modifications of the previously described switching cost model based on travel time, and present results in Table 5.2. A constant model with cost of one irrespective of location is used to investigate the effectiveness of location-dependent switching costs. A normalized model, calculated by dividing the travel-derived switching matrix by its mean, investigates the importance of scaling the cost to the Bernoulli reward. In contrast, an inflated model has the switching matrix scaled up ten-fold. Improvement was comparable across all cost models considered. These results show that the MABSC performs consistently well with a range of switching cost models and probability spreads.

[1] Computed with Matlab R2012b on Windows 7 (64bit), Intel i5-3450, 16GB of RAM

Table 5.1: Computation Time by Lookahead Horizon and Field of Probabilities

Horizon	Narrow Range	Wide Range	Time (s)
ϵ-g	41.71	156.74	-
ϵ-d	54.14	174.96	-
MAB	60.97	183.95	-
1	63.46	183.82	1.3e-04
2	65.07	187.11	1.8e-03
3	65.97	190.13	1.7e-02
5	67.23	190.43	1.3e-00
7	67.34	190.40	1.0e+02

Table 5.2: Percentage Reward Rate Improvement Over Touring Survey

Switching Cost	Narrow Range	Wide Range
Constant	64.78	188.80
Travel Time	65.97	190.13
Normalized	64.02	187.05
Inflated	68.35	188.37

Chapter 6

Conclusion

We have shown that a multi-armed bandit formulation for adaptive acoustic relay positioning can address the direct but poorly-known coupling of channel properties to the relay's physical location. In particular, the MAB algorithm optimally trades off real-time exploration of the field with exploitation of high-performing sites, and without any tuning parameters. We augmented the canonical MAB to include switching costs based on proven properties of the optimal solution, and demonstrated in the field and using synthetic data that the MABSC scheme can achieve at least twice the throughput of a roughly-tuned greedy method, and is never beaten. As well, the field statistics are more quickly and accurately discovered by multi-armed bandit decisions; the baseline MAB gathers global information most efficiently while the MABSC is efficient at both information-gathering and maximizing cumulative reward.

We justify our assumption of stationarity over the course of the mission using transmission statistics of the acoustic channel. However, we have observed significant day-to-day spatial variation in our field experiments, most likely tied to changing weather conditions. Longer missions and rapidly-changing weather conditions may weaken this assumption. This variability may be properly addressed by a restless bandit reformulation, though no optimal policy like Gittins indices has been developed. The multi-armed bandit algorithm may also be re-initialized periodically, according to the time scale of the mission. Further, the multi-armed bandit requires discretization of the space in order to facilitate learning and decision-making. The selection of potential relay locations may be dependent on the mission specifications and the spatial variability in the acoustic channel. This problem may be reformulated as an arm-acquiring bandit, in which new arms are spawned in high-performing regions or regions with high variability, to provide additional resolution of the field. Since the computational cost of enumeration scales exponentially with the number of arms, the relay must choose from a relatively small space of potential locations. As future work, we plan to incorporate multiple relays in a three-hop transmission link, where switching costs are dependent on the location and state of each relay.

Appendix A

Figures

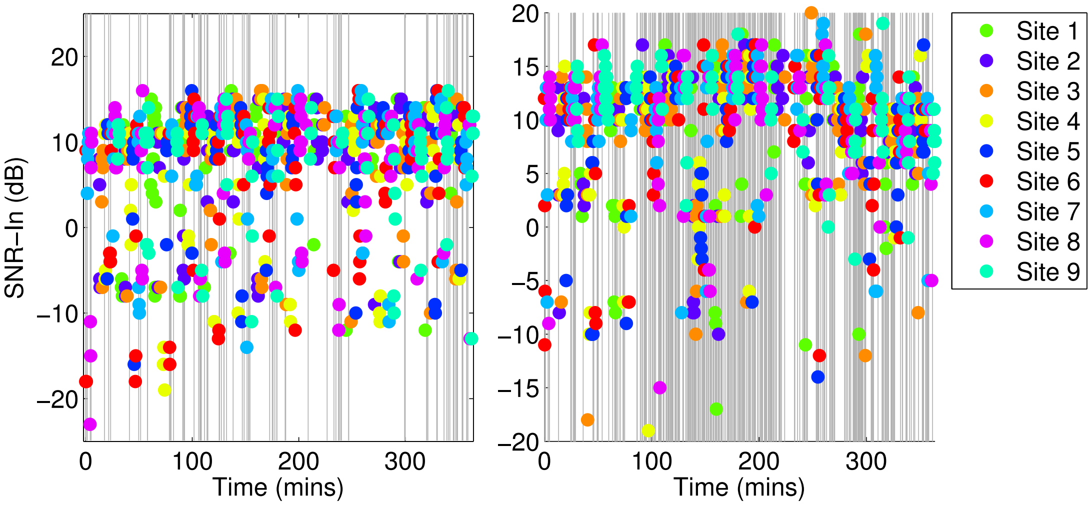


Figure A-1: SNR-In values against mission time for HybridSetB Source-to-Relay (left) and Relay-to-End (right). Grey lines represent lost packets. Sites are visited in the same order.

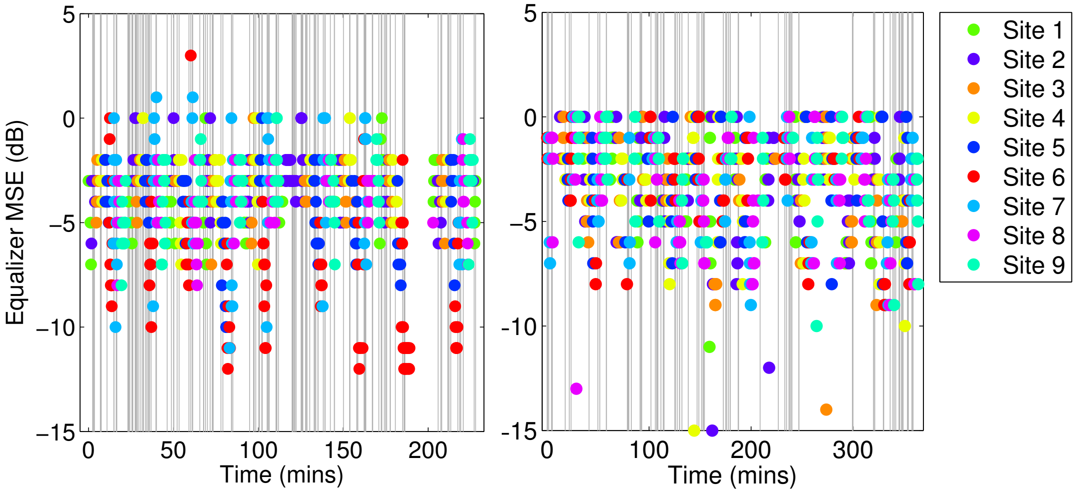


Figure A-2: Mean-squared-error (MSE) out of the Equalizer against mission time for HybridSetA (left) and B (right). Grey lines represent lost packets. Sites are visited in the same order and data is for source-to-relay transmission.

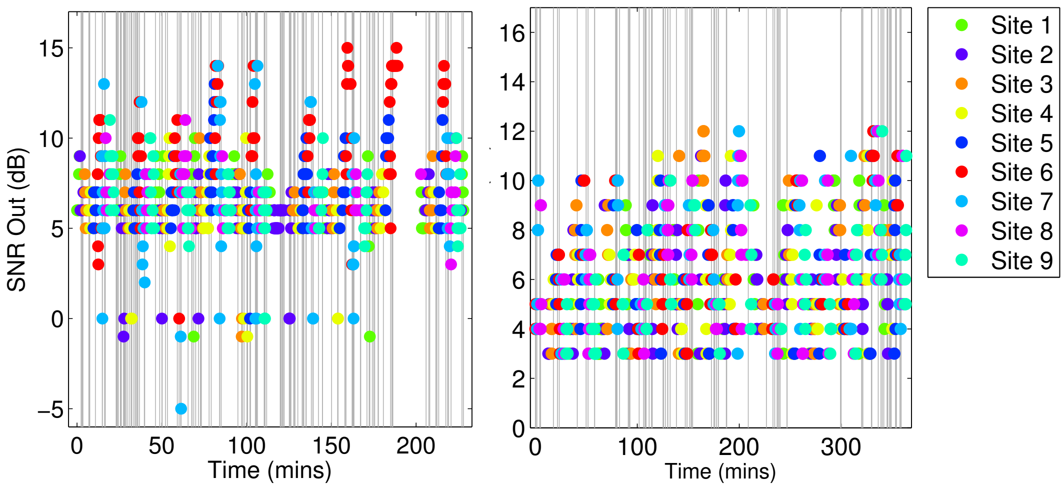


Figure A-3: SNR-Out values against mission time for HybridSetA (left) and B (right). Grey lines represent lost packets. Sites are visited in the same order and data is for source-to-relay transmission.

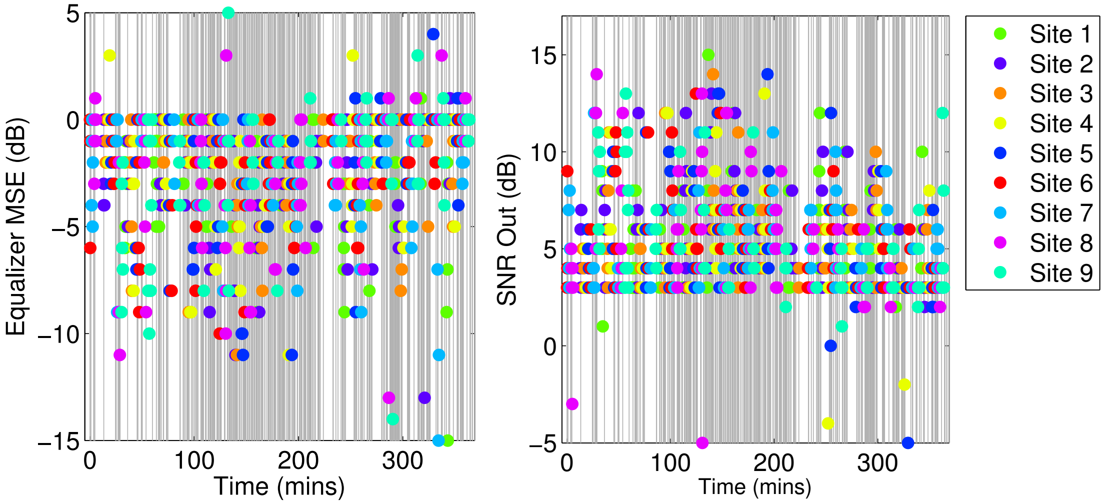


Figure A-4: Mean-squared-error (MSE) out of the Equalizer (left) and SNR-Out values (right) for HybridSetB Relay-to-End transmission. Grey lines represent lost packets. Sites are visited in the same order.

Appendix B

Tables

Table B.1: Autonomous Kayaks Components and Dimensions

Component	Part
Hull	WaveSpot Fuse 35 Kids Whitewater Kayak
Thruster	Minn-Kota Riptide 55 Trolling Motor
Motor Driver	Roboteq LDC1450 Brushed DC Motor Controller
Servo	ServoCity MS530-1 Mega Servo
Batteries	Lithium Iron Phosphate (LiFePO4) 12.8V 60Ah and 100Ah
Compass	Ocean Server OS5000
RF Comms	Ubiquiti 2.4GHz Bullet M
Freewave Comms	Freewave FGR2-PE 900 MHz
Radio Control	Futaba 7 Channel 2.5GHz Controller

Web Reference: `https://wikis.mit.edu/confluence/display/hovergroup/Home`

Table B.2: Micro-Modem Transmission Rates

Rate	ECC	Modulation Type	Bytes per Frame	Maximum Frames
0	Conv(2,1,9)	FH-FSK	32	1
1*	BCH (128:8)	QPSK	64	3
2*	DSSS-15	QPSK	64	3
3	DSS-7	QPSK	256	2
4	BCH(64:10)	QPSK	256	2
5	Hamming(14:9)	QPSK	256	8
6	DSS-15	QPSK	32	6

FH-FSK: Frequency-hopped frequency-shift-keyed

QPSK: Quadrature phase-shift-keyed.

*These rates were used in field experiments.

Source: `http://acomms.whoi.edu/umodem/documentation.html`

Appendix C

Code

Definitions:

n	number of arms
m	horizon length
a	number of successes
b	number of failures
β	discount factor
β_i	intermediate discount factor
β_e	expanded discount factor
$C(i,j)$	switching cost matrix
$\bar{c}$	switching cost submatrix (vector)
$\bar{p}$	policy indices
$\bar{p}_i$	intermediate indices
$\bar{p}_e$	expanded indices
$\bar{\mu}$	packet success rates

Algorithm 1 Policy Index Enumeration

```
Require: $m > 1$
  for $i = m$ to 2 increment by $-1$ do
    for $j = 0$ to $i$ do
      $\beta_i \leftarrow \beta_i + \beta^j$
    end for
    $\beta_e \leftarrow \beta^m / \beta_i$
    for $k = n$ to $n^m$ increment by $n$ do
      $\beta_i \leftarrow \beta_i + \beta^j$
      if $k/n \bmod n > 0$ then
        $\bar{c} \leftarrow C(k/n \bmod n, :)$
      else
        $\bar{c} \leftarrow C(n, :)$
      end if
      $\bar{p}_e(k - (n-1) : k) \leftarrow \bar{p}_e(k - (n-1) : k) - \bar{c}^T * \beta_e$
      $\bar{p}_e(k - (n-1) : k) \leftarrow \bar{p}_e(k - (n-1) : k) + \bar{\mu}^i * \beta_e$
      if $i = m$ then
        for $x = 1$ to $n$ do
          $\bar{p}_e(k - (n-1) : k) \leftarrow \bar{p}_e(k - (n-1) : k) - \bar{\mu}^i * \text{GILookup}(a + i, b)$
        end for
      end if
      if $m > 2$ then
        $\bar{p}_i(k/n) = \max \bar{p}_e(k - (n-1) : k)$
      else
        $\bar{p}(k/n) = \bar{p}(k/n) + \max \bar{p}_e(k - (n-1) : k)$
      end if
    end for
  end for
```

Bibliography

[1] R. Agrawal, M.V. Hedge, and D. Teneketzis. Asymptotically efficient adaptive allocation rules for the multiarmed bandit problem with switching cost. *IEEE Transactions on Automatic Control*, 33(10):899–906, 1988.

[2] I. F. Akyildiz, D. Pompili, and T. Melodia. Underwater acoustic sensor networks: research challenges. *Ad hoc networks*, 3(3):257–279, 2005.

[3] I. F. Akyildiz, D. Pompili, and T. Melodia. State-of-the-art in protocol research for underwater acoustic sensor networks. In *Proc. 1st ACM International Workshop on Underwater Networks*, pages 7–16, 2006.

[4] A Alvarez, A Caffaz, A Caiti, G. Casalino, L Gualdesi, A. Turetta, and R. Viviani. Folaga: a low-cost autonomous underwater vehicle combining glider and auv capabilities. *Ocean Engineering*, 36(1):24–38, 2009.

[5] M. Asawa and D. Teneketzis. Multi-armed Bandits with Switching Penalties. *IEEE Transactions on Automatic Control*, 41(3):328–348, 1996.

[6] W. W.L. Au and K. Banks. The acoustics of the snapping shrimp synalpheus parneomeris in kaneohe bay. *The Journal of the Acoustical Society of America*, 103:41, 1998.

[7] J. S. Banks and R. K. Sundaram. Switching Costs and the Gittins Index. *Econometrica*, 62(3):687–694, 1994.

[8] M. R. Benjamin, J. J. Leonard, H. Schmidt, and P. M. Newman. An overview of moos-ivp and a brief users guide to the ivp helm autonomy software. Technical report, Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory (MIT-CSAIL), 2009.

[9] D. Bertsimas and J. Niño Mora. Restless Bandits, Linear Programming Relaxations, and a Primal-Dual Index Heuristic. *Operations Research*, 48(1):80–90, 2000.

[10] J. B. Bowlin, J. L. Spiesberger, T. F. Duda, and L. E. Freitag. Ocean acoustical ray-tracing : Software Ray. Technical report, WHOAS at MBLWHOI Library, 1992.

[11] L. Brun. Rov/auv trends: Market and technology. *Marine Technology Reporter*, pages 48–51, 2012.

[12] A. Cerpa and D. Estrin. Ascent: Adaptive self-configuring sensor networks topologies. *mobile computing, IEEE transactions on*, 3(3):272–285, 2004.

[13] M. Cheung, J. Leighton, and F. Hover. Autonomous Mobile Acoustic Relay Positioning as a Multi-Armed Bandit with Switching Costs. In *Proc. IEEE International Conference on Intelligent Robots and Systems (IROS), to appear*, 2013.

[14] M. Cheung, J. Leighton, and F. Hover. Field experiments in acoustic relay positioning as a multi-armed bandit with switching costs. In *International Symposium on Robotics Research (ISRR), Submitted To*, 2013.

[15] M. Cheung, J. Leighton, and F. Hover. Multi-armed Bandit Formulation for Autonomous Mobile Acoustic Relay Adaptive Positioning. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[16] M. Chitre, S. Shahabudeen, and M. Stojanovic. Underwater Acoustic Communications and Networking: Recent Advances and Future Challenges. *Marine Technology Society Journal*, 42(1):103–116, 2008.

[17] J. Curcio, T. Schneider, M. Benjamin, and A. Patrikalakis. Autonomous surface craft provide flexibility to remote adaptive oceanographic sampling and modeling. In *OCEANS 2008*, pages 1–7. IEEE, 2008.

[18] C. Detweiler, M. Doniec, I. Vasilescu, E. Basha, and D. Rus. Autonomous Depth Adjustment for Underwater Sensor Networks. In *Proc. 5th ACM International Workshop on Underwater Networks*, pages 12:1–12:4, 2010.

[19] F. Dusonchet and M.O. Hongler. Optimal hysteresis for a class of deterministic deteriorating two-armed bandit problem with switching costs. *Automatica*, 39(11):1947–1955, 2003.

[20] B. Englot and F. Hover. Inspection planning for sensor coverage of 3d marine structures. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 4412–4417. IEEE, 2010.

[21] R. Eustice, H. Singh, J. J. Leonard, M. Walter, and R. Ballard. Visually navigating the rms titanic with slam information filters. In *Robotics: Science and Systems*, pages 57–64, 2005.

[22] L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, and K. Ball. The WHOI Micro-Modem: An Acoustic Communications and Navigation System for Multiple Platforms. In *Proc. MTS/IEEE OCEANS*, volume 2, pages 1086–1092, Sept. 2005.

[23] Y. Gai, B. Krishnamachari, and R. Jain. Learning Multiuser Channel Allocations in Cognitive Radio Networks: A Combinatorial Multi-Armed Bandit Formulation. In *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum*, pages 1–9, Apr. 2010.

[24] J. C. Gittins. Bandit Processes and Dynamic Allocation Indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):148–177, 1979.

[25] J. C. Gittins, R. Weber, and K. D. Glazebrook. *Multi-armed Bandit Allocation Indices*, volume 25. Wiley Online Library, 1989.

[26] Z. Guo, G. Colombo, B. Wang, J. Cui, D. Maggiorini, and G. Rossi. Adaptive routing in underwater delay/disruption tolerant sensor networks. In *Wireless on Demand Network Systems and Services, 2008. WONS 2008. Fifth Annual Conference on*, pages 31–39. IEEE, 2008.

[27] T. Jun. A survey on the bandit problem with switching costs. *De Economist*, 152(4):513–541, 2004.

[28] V. Krishnamurthy and R. J. Evans. Hidden Markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking. *IEEE Transactions on Signal Processing*, 49(12):2893–2908, Dec. 2001.

[29] O.B. Kroemer, R. Detry, J. Piater, and J. Peters. Combining active learning and reactive control for robot grasping. *Robotics and Autonomous Systems*, 58(9):1105–1116, 2010.

[30] C. Kunz, C. Murphy, R. Camilli, H. Singh, J. Bailey, R. Eustice, M. Jakuba, K. Nakamura, C. Roman, T. Sato, et al. Deep sea underwater robotic exploration in the ice-covered arctic ocean with auvs. In *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pages 3654–3660. IEEE, 2008.

[31] J. Le Ny, M. Dahleh, and E. Feron. Multi-Agent Task Assignment in the Bandit Framework. In *Proc. 45th IEEE Conference on Decision and Control*, pages 5281–5286, Dec. 2006.

[32] J. Le Ny and E. Feron. Restless bandits with switching costs: Linear programming relaxations, performance bounds and limited lookahead policies. In *Proc. IEEE American Control Conference*, pages 6–12, 2006.

[33] N. E. Leonard, D. A. Paley, R. E. Davis, D. M. Fratantoni, F. Lekien, and F. Zhang. Coordinated control of an underwater glider fleet in an adaptive ocean sampling field experiment in monterey bay. *Journal of Field Robotics*, 27(6):718–740, 2010.

[34] A. Munafò, E. Simetti, A. Turetta, A. Caiti, and G. Casalino. Autonomous underwater vehicle teams for adaptive ocean sampling: a data-driven approach. *Ocean Dynamics*, 61(11):1981–1994, 2011.

[35] J. Niño-Mora. A (2/3)n3 Fast-Pivoting Algorithm for the Gittins Index and Optimal Stopping of a Markov Chain. *INFORMS J. on Computing*, 19(4):596–606, Oct. 2007.

[36] J. Niño-Mora. A faster index algorithm and a computational study for bandits with switching costs. *INFORMS J. on Computing*, 20(2):255–269, 2008.

[37] A. Pascoal, P. Ridao, A. Birk, M. Eichhorn, L. Brignone, M. Caccia, J. Alvez, R.S. Santos, and J. Kalwa. Marine robotic systems of self-organizing, logically linked physical nodes. Annex 1, EC Project ICT 288704, 2011.

[38] M. B. Porter and H. P. Bucker. Gaussian beam tracing for computing ocean acoustic fields. *The Journal of the Acoustical Society of America*, 82:1349, 1987.

[39] W. B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. Wiley-Blackwell, 2007.

[40] J. C. Preisig. Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment. *The Journal of the Acoustical Society of America*, 118(1):263–278, 2005.

[41] F. Radlinski, R. Kleinberg, and T. Joachims. Learning diverse rankings with multi-armed bandits. In *Proc. 25th International Conference on Machine Learning (ICML)*, pages 784–791, 2008.

[42] T. Schneider and H. Schmidt. Model-based adaptive behavior framework for optimal acoustic communication and sensing by marine robots. *IEEE Journal of Oceanographic Engineering*, 38:522–533, 2013.

[43] S. Shankar and Chitre. Tuning an underwater communication link. In *OCEANS*, 2013.

[44] I. M. Sonin. A generalized Gittins index for a Markov chain and its recursive calculation. *Statistics and Probability Letters*, 78(12):1526–1533, 2008.

[45] M. Stojanovic. Recent advances in high-speed underwater acoustic communications. *Oceanic Engineering, IEEE Journal of*, 21(2):125–136, 1996.

[46] M. Stojanovic and J. Preisig. Underwater acoustic communication channels: Propagation models and statistical characterization. *Communications Magazine, IEEE*, 47(1):84–89, Jan. 2009.

[47] P. Stone and S. Kraus. To Teach or not to Teach?: Decision making under uncertainty in ad hoc teams. In *Proc. 9th International Conference on Autonomous Agents and Multiagent Systems*, volume 1, pages 117–124, 2010.

[48] T. Takasu and A. Yasuda. Development of the low-cost rtk-gps receiver with an open source program package rtklib. In *International Symposium on GPS/GNSS*, 2009.

[49] Y.F. Wei, F. Yu, M. Song, and Y. Zhang. Transmission control protocol throughput optimisation in cooperative relaying networks through relay selection. *IET Communications*, 5(16):2257–2265, 2011.

[50] L. Whitcomb. Underwater robotics: Out of the research laboratory and into the field. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 1, pages 709–716. IEEE, 2000.

[51] L. Whitcomb, D. Yoerger, H. Singh, and D. Mindell. Towards precision robotic maneuvering, survey, and manipulation in unstructured undersea environments. In *Robotics Research*, pages 45–54. Springer, 1998.

[52] P. Whittle. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 25:287–298, 1988.

[53] N. K. Yilmaz, C. Evangelinos, P. Lermusiaux, and N. M. Patrikalakis. Path planning of autonomous underwater vehicles for adaptive sampling using mixed integer linear programming. *Oceanic Engineering, IEEE Journal of*, 33(4):522–537, 2008.